

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 16:43:02 ; Search time 44 Seconds
(without alignments)
1248.168 Million cell updates/sec

Title: PCT-US03-21094-2

Perfect score: 1814

Sequence: 1 MLFQQAWLRQKLLVLSIA.....FFHPNQKPYQITGRLNWP 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1646	90.7	321	23	AAO21520 Human drug metabol
2	1646	90.7	340	23	AAE14442 Human drug metabol
3	935	51.5	175	23	ABE04738 Human heparan sulf
4	780	43.0	307	20	AAI17063 Human 3-OST-1 prot
5	780	43.0	307	24	AAO26606 Human 3-OST-1 prot
6	780	43.0	321	23	ABP41427 Human ovarian anti
7	772.5	42.6	255	23	AAO21521 255-mer human drug
8	758	41.8	307	22	ABE30928 Peptide #3579 enco
9	758	41.8	307	22	ABE36116 Peptide #3622 enco

10	758	41.8	307	22	ABB21504 Protein #3503 enco
11	758	41.8	307	22	AAE56899 Human brain expres
12	758	41.8	307	22	AAE59287 Human bone marrow
13	758	41.8	307	22	AAE29607 Peptide #3644 enco
14	758	41.8	307	23	ABG38901 Human peptid enco
15	757.5	41.8	311	20	AAI17062 Mouse 3-OST-1 prot
16	646.5	35.6	384	22	ABE63532 Drosophila melanog
17	642.5	35.4	456	22	AAI17067 Human 3-OST-4 prot
18	642.5	35.4	456	24	ABB82784 Human 3OST4 polype
19	637.5	35.1	406	20	AAI17065 Human 3-OST-3A pro
20	636.5	35.1	390	20	AAI17066 Human 3-OST-3B pro
21	636.5	35.1	390	22	AAE95507 Human protein sequ
22	635.5	35.0	367	20	AAI17064 Human 3-OST-2 prot
23	635.5	35.0	367	22	AAU12275 Human PRO5004 poly
24	635.5	35.0	367	24	ABU66673 Human PRO polypept
25	635.5	35.0	367	24	ABU66949 Human secreted/tra
26	635.5	35.0	367	24	ABU59754 Novel secreted and
27	627.5	34.6	391	22	AAE70115 Human 3-OST-3B. H
28	618.5	34.1	271	22	AAE70116 Human 3-OST-3A. H
29	606.5	33.4	391	22	AAE70114 Murine 3-OST-3B.
30	581.5	32.1	291	20	AAI17070 C. elegans putativ
31	562	31.0	405	23	ABU65174 Human NOV93a prote
32	435.5	24.0	298	22	ABE70128 Drosophila melanog
33	333	18.4	286	20	AAI17069 Human NST-2 protei
34	330.5	18.2	1048	22	ABE63796 Drosophila melanog
35	319.5	17.6	873	23	AAE25018 Human drug metabol
36	318.5	17.6	882	15	AAE43959 Human heparan sulphate
37	316.5	17.4	882	23	ABG31772 Rat N-acetylglucos
38	312.5	17.2	882	23	ABG31773 Human N-acetylgluc
39	311	17.1	284	20	AAI17068 Human NST-1 protei
40	282	15.5	240	22	ABE70127 Drosophila melanog
41	229.5	12.7	123	22	ABE16656 Human nervous syst
42	177	9.8	274	22	AAE74729 Human colon cancer
43	171	9.4	72	22	AAU21387 Human novel foetal
44	162	8.9	244	24	ABE32785 Human 3OST4 polype
45	160	8.8	157	22	ABG55384 Human liver peptid

ALIGNMENTS

RESULT 1
AAO21520
ID AAO21520 standard; Protein; 321 AA.
XX AAO21520;
AC AAO21520;
XX
DT 16-AUG-2002 (first entry)
XX
DE Human drug metabolising enzyme protein.
XX
KW Human drug-metabolising protein; pharmacogenomic analysis;
KW sulfotransferase drug-metabolising enzyme; gene therapy.
XX
OS Homo sapiens.

Key	Location/Qualifiers
FT Region	2..4
FT Region	/note= "Protein kinase C phosphorylation site"
FT Region	23..28
FT Region	/note= "N-myristoylation site"
FT Region	27..30
FT Modified-site	/note= "Casein kinase II phosphorylation site"
FT Region	50..53
FT Region	/note= "Asn is N-glycosylated"
FT Region	72..77
FT Region	/note= "N-myristoylation site"
FT Region	76..81
FT Modified-site	/note= "N-myristoylation site"
FT Region	148..151
FT Region	/note= "Asn is N-glycosylated"
FT Region	150..152
FT Region	/note= "Protein kinase C phosphorylation site"

FT Region 161..163 /note= "Protein kinase C phosphorylation site"
FT Modified-site 179..182 /note= "Asn is N-glycosylated"
FT Modified-site 262..265 /note= "Asn is N-glycosylated"
FT Region 314..316 /note= "Protein kinase C phosphorylation site"
XX WO200242437-A2.
XX 30-MAY-2002.
XX 16-NOV-2001; 2001WO-US43109.
XX 27-NOV-2000; 2000US-252895P.
XX 14-DEC-2000; 2000US-0735935.
XX (PEKE) PE CORP NY.
XX Guegler K, Ketchum KA, Di Francesco V, Beasley EM;
XX WPI; 2002-479949/51.
XX N-PSDB; AAL38608, AAL38609.
XX New human drug-metabolizing enzyme, related to the sulfotransferase
FT drug-metabolizing enzyme subfamily, useful as a model for developing
FT human therapeutic targets and as a target for human therapeutics -
XX Claim 1; Fig 2; 65pp; English.
XX The invention relates to an isolated human drug-metabolising polypeptide
CC which is related to the sulfotransferase drug-metabolising enzyme
CC subfamily, consisting of comprising of a sequence of 321 amino acids,
CC given in the specification, or its fragment comprising 10 contiguous
CC amino acids, or an amino acid sequence of an allelic variant or ortholog.
CC The protein of the invention, its allelic variant or ortholog or
CC fragment, is used to identify a modulator of a human drug-metabolising
CC enzyme protein. The drug-metabolising enzyme proteins isolated from
CC humans and their human/mammalian orthologues serve as targets for
CC identifying agents for use in mammalian therapeutic applications, and
CC biological assays related to the drug-metabolising enzyme proteins. The
CC proteins can also be used in screening assays to screen a compound for
CC its ability to stimulate or inhibit interaction between drug-metabolising
CC enzyme proteins and a molecule that normally interacts with the drug-
CC metabolising enzyme proteins. The proteins also provide a target for
CC diagnosing a disease or predisposition to disease mediated by the
CC peptide, and in pharmacogenomic analysis. The nucleic acid encoding the
CC protein of the invention also provides vectors for gene therapy in
CC patients with aberrant expression of a gene encoding the drug-
CC metabolising enzyme. This sequence represents the human drug metabolising
CC enzyme protein of the invention.
XX SQ Sequence 321 AA;
Query Match 90.7%; Score 1646; DB 23; Length 321;
Best Local Similarity 100.0%; Pred. No. 2e-155;
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 36 RLQPICTIEGLGARTQAEFFLRALQFKRGLLHFRKGNASKQVRLHDLVQQLPKAI 95
DB 11 RLQPICTIEGLGARTQAEFFLRALQFKRGLLHFRKGNASKQVRLHDLVQQLPKAI 70
QY 96 IGVKGGTRALLEMLNHPAVKASQEIHFDDNENYKGIWYRKMPFSYQQITIEK 155
DB 71 IGVKGGTRALLEMLNHPAVKASQEIHFDDNENYKGIWYRKMPFSYQQITIEK 130
QY 156 SPAYFTTEEPRIYKNSIKLLIVREPTTAISDYQVLGKGRKNTYKFKLAI 215
DB 131 SPAYFTTEEPRIYKNSIKLLIVREPTTAISDYQVLGKGRKNTYKFKLAI 190
QY 216 DPNTCEVNTKYKAVRTSYTKHLERWLKYFPIEQFHVVDGLITEPLQLVEKFLNL 275

Db 191 DPNTCEVNTKYKAVRTSYTKHLERWLKYFPIEQFHVVDGLITEPLQLVEKFLNL 250
QY 276 PPRISQYNLYFNATRGFYCLRNIIFNKCLAGSKGRIHPEVDPSPVITTKLKRFHPNQKF 335
Db 251 PPRISQYNLYFNATRGFYCLRNIIFNKCLAGSKGRIHPEVDPSPVITTKLKRFHPNQKF 310
QY 336 YQITGRTLNWP 346
Db 311 YQITGRTLNWP 321
RESULT 2
ID AAE14442 standard; Protein; 340 AA.
XX AAE14442;
XX AC AAE14442;
XX 26-MAR-2002 (first entry)
XX Human drug metabolising enzyme (DME)-5.
XX Human; drug metabolising enzyme; DME-5; autoimmune; inflammatory;
KW cell proliferative; developmental; endocrine; eye; metabolic; AIDS;
KW gastrointestinal disorder; liver disorder; cancer; arteriosclerosis;
KW adult respiratory distress syndrome; anaemia; epilepsy; hypothyroidism;
KW hypothyroidism; pituitary; diabetes; hypogonadism; conjunctivitis;
KW glaucoma; cystic fibrosis; hypercholesterolaemia; gastritis;
KW peptic ulcer; hepatitis; gene therapy.
XX OS Homo sapiens.
XX WO200190334-A2.
XX 29-NOV-2001.
XX 25-MAY-2001; 2001WO-US17150.
XX 25-MAY-2000; 2000US-207901P.
PR 01-JUN-2000; 2000US-208983P.
PR 07-JUN-2000; 2000US-209861P.
PR 15-JUN-2000; 2000US-211825P.
PR 22-JUN-2000; 2000US-213744P.
XX (INCY-) INCYTE GENOMICS INC.
XX Yue H, Sanjanwala MS, Baughn MR, Gandhi AR, Ring HZ, Elliott V;
PI Walla NK, Yang J, Khan FA, Ramkumar J, Tang YT, Hafalia A, Lal P;
PI Nguyen DB, Yao WG, Lee EA, Tribouley CM, Patterson C, Lu Y;
PI Burford N, Ding L, Bruns CM, Kearney L, Reddy R;
XX WPI; 2002-097650/13.
DR N-PSDB; AAD24010.
XX New human drug metabolizing enzymes and polynucleotides encoding the
PT enzyme for diagnosing, preventing or treating cell proliferative,
PT autoimmune/inflammatory, endocrine, eye, metabolic and gastrointestinal
PT disorders -
XX Claim 1; Page 140-141; 158pp; English.
XX The present sequence is human drug metabolising enzyme (DME)-5.
CC DME polypeptide, polynucleotide and modulators are useful for
CC diagnosis, treatment and prevention of autoimmune/inflammatory,
CC cell proliferative, developmental, endocrine, eye, metabolic,
CC and gastrointestinal disorders, including liver disorders.
CC The autoimmune/inflammatory disorders treatable include
CC AIDS, adult respiratory distress syndrome, Addison's disease,
CC allergies, anaemia, asthma, atherosclerosis, osteoporosis, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, Crohn's disease, atopic
CC dermatitis, diabetic mellitus, Graves' disease, glomerulonephritis,
CC rheumatoid arthritis, scleroderma, systemic lupus erythematosus,
CC systemic sclerosis, ulcerative colitis, haemodialysis and uveitis,
CC viral, bacterial, fungal, parasitic, protozoal, helminthic infections

QY 256 DRLITEPLDELQVEXKFLNPPRIISQVNLVFNATRGFYCLRENIIFNKCLAGSKGRIHPE 315
 DB 218 DRLIRDFPEIQVERFLKSLPOINASNFYFNKTKGYCLR-DSGRDCLHESKGRAPQ 276
 QY 316 VDPSTVITKLRKFFPHFPNQKFYQITGRTLW 345
 DB 277 VDPKLNKLHXYFHEFNKFFELVGRITFDW 306
 RESULT 6
 ABP41427
 ID ABP41427 standard; Protein; 321 AA.
 AC ABP41427;
 XX
 DT 22-AUG-2002 (first entry)
 XX
 DE Human ovarian antigen HHP5H76, SEQ ID NO:2559.
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;
 KW inflammatory condition; immune disorder; blood disorder;
 KW cardiovascular disorder; respiratory disorder; neurological disorder;
 KW gastrointestinal disorder; urinary system disorder; drug screening;
 KW gene therapy; chromosome mapping; forensic analysis;
 KW antibody preparation; cytotoxic; immunomodulatory; neuroprotective;
 KW antiinflammatory; gynaecological; reproductive.
 XX
 OS Homo sapiens.
 XX
 PN WO200200677-A1.
 XX
 PD 03-JAN-2002.
 XX
 PF 07-JUN-2001; 2001WO-US18569.
 XX
 PR 07-JUN-2000; 2000US-209467P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;
 DR WPI; 2002-147878/19.
 DR N-PSDB; ABQ54504.
 XX
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,
 PT useful in the prevention, treatment and diagnosis of cancer (e.g.
 PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11; SEQ ID No 2559; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovary and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and

CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 321 AA;
 Query Match 43.0%; Score 780; DB 23; Length 321;
 Best Local Similarity 46.4%; Pred. No. 4.8e-69;
 Matches 153; Conservative 64; Mismatches 89; Indels 24; Gaps 6;
 QY 16 LGSIAVGSLLIVAVRGSLDLRLOPCPIEGRILGGARTQAEFFPLRALQFKRGLLHFRKGN 75
 DB 15 MAALLGAVL-LVAQ-----PQLVPSRP-----AELQQELLRKAGTLQD----- 53
 QY 76 ASKEQVRLHDLVQQLPKAIIIGVRKGGTRALLEMLNLHPAVVAKASQEIHFDFDNDENYKG 135
 DB 54 DVRDGVAPNGSAQQLPQTIIIGVRKGGTRALLEMLSLHDPVAAAEVHFFDWEHYSHG 113
 QY 136 IEWYKRPFPSPYPOQITTEKSPAYITSEVPRIYKNSIKLLIIVREPTTTRAIADYQ 195
 DB 114 LGWILSQMPFSPHQITVEKTPAYTSPKPVSVYMNPSIRLLLLRDPSEKVLSDYQ 173
 QY 196 VLEGKSKNKTYKFEKLAIDNTCEVNTKYKAVRTSIYTKLERWLKYPPIEQPHWDG 255
 DB 174 VFYEMQKHKYPSEIEFLVRDG--RLNVDYKALNRSLYHVENQWLRFFPLRHIVDG 231
 QY 256 DRLITEPLDELQVEXKFLNPPRIISQVNLVFNATRGFYCLRENIIFNKCLAGSKGRIHPE 315
 DB 232 DRLIRDFPEIQVERFLKSLPOINASNFYFNKTKGYCLR-DSGRDCLHESKGRAPQ 290
 QY 316 VDPSTVITKLRKFFPHFPNQKFYQITGRTLW 345
 DB 291 VDPKLNKLHXYFHEFNKFFELVGRITFDW 320
 RESULT 7
 AA021521
 ID AA021521 standard; Protein; 255 AA.
 XX
 AC AA021521;
 XX
 DT 16-AUG-2002 (first entry)
 XX
 DE 255-mer human drug metabolising enzyme related protein.
 XX
 KW Human drug-metabolising protein; pharmacogenomic analysis;
 KW sulfotransferase drug-metabolising enzyme; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200242437-A2.
 XX
 PD 30-MAY-2002.
 XX
 PF 16-NOV-2001; 2001WO-US43109.
 XX
 PR 27-NOV-2000; 2000US-252895P.
 PR 14-DEC-2000; 2000US-0735935.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Guegler K, Ketchum KA, Di Francesco V, Beasley EM;
 XX WPI; 2002-479949/51.
 XX
 PT New human drug-metabolizing enzyme, related to the sulfotransferase

PT drug-metabolizing enzyme subfamily, useful as a model for developing
PT human therapeutic targets and as a target for human therapeutics -
XX
PS Disclosure; Fig 2; 65pp; English.
XX
CC The invention relates to an isolated human drug-metabolising polypeptide
CC which is related to the sulfoxtransferase drug-metabolising enzyme
CC subfamily, consisting of or comprising a sequence of 321 amino acids,
CC given in the specification, or its fragment comprising 10 contiguous
CC amino acids, or an amino acid sequence of an allelic variant or ortholog.
CC The protein of the invention, its allelic variant or ortholog or
CC fragment, is used to identify a modulator of a human drug-metabolising
CC enzyme protein. The drug-metabolising enzyme proteins isolated from
CC humans and their human/mammalian orthologues serve as targets for
CC identifying agents for use in mammalian therapeutic applications, and
CC biological assays related to the drug-metabolising enzyme proteins. The
CC proteins can also be used in screening assays to screen a compound for
CC its ability to stimulate or inhibit interaction between drug-metabolising
CC enzyme proteins and a molecule that normally interacts with the drug-
CC metabolising enzyme proteins. The proteins also provide a target for
CC diagnosing a disease or predisposition to disease mediated by the
CC peptide, and in pharmacogenomic analysis. The nucleic acid encoding the
CC protein of the invention also provides vectors for gene therapy in
CC patients with aberrant expression of a gene encoding the drug-
CC metabolising enzyme. This sequence represents a 255-mer human protein
CC related to the human drug-metabolising enzyme protein of the invention.
XX
SQ Sequence 255 AA;
Query Match 42.6%; Score 772.5; DB 23; Length 255;
Best Local Similarity 53.1%; Pred. No. 1.9e-68;
Matches 137; Conservative 52; Mismatches 66; Indels 3; Gaps 2;
QY 88 QQLPKAIIIGVRKGGTRALLEMLNHPAVVYKASQEIHFDDNENYKGIWYRKKMPFSY 147
Db 1 QQLPQTIIIGVRKGGTRALLEMLSLHPDVAANEVHFFDWEHSHGLWYLSQMPFSW 60
QY 148 PQQITTEKSPAYITEVPRYIKYNSIKLLIIVREPTTRAIISDYTVLGEKRNKTY 207
Db 61 PHQITVEKTPAYFTSPKVPERVYSMPNSIRLLILRLDPSERVLSDYTVQVYNHMQHKPY 120
QY 208 YKPEKLAIDNTECVNTKTKAVRTSIYTKHLEBWLKVFTEQVHVVDGRLITEPELQ 267
Db 121 PSIEEFLVRDG--RLNVVDYKALNRSLYHVHMQNWLRFPPFLRHHIVDGDLEIDPPEIQ 178
QY 268 LVEKFLNLPRIISQYNLYFNATRGFYCLRENIFNFKLAGSKGRIHPVDPSPVITKIKKF 327
Db 179 KVERFLKSPQINASNFYFNKTKGYCLR--DSGRDCLHESKGRAHPQVDPKLLINKHEY 237
QY 328 FHPFNOKFYQITORTLNW 345
Db 238 FHEPNKFFELVGRTFDW 255
RESULT 8
ABB30928
ID ABB30928 standard; Peptide; 307 AA.
AC ABB30928;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #3579 encoded by breast cell single exon nucleic acid probe.
XX Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX Homo sapiens.
OS
XX WO200157271-A2.
PN
XX 09-AUG-2001.
PD
XX

PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 27; SEQ ID NO 13896; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 307 AA;
Query Match 41.8%; Score 758; DB 22; Length 307;
Best Local Similarity 46.2%; Pred. No. 7e-67;
Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;
QY 16 LGSLLVGLLYLVARVGSIDRLQIPICPIEGRGGARTQAEFFLRALQFKRGLLHFRKGN 75
Db 7 MAALLIGAVL-LVAQ----PQLVPSRP-----AELGQQLLRKAGTLQD----- 45
QY 76 ASKEQVRLHDLVQOLPKAIIIGVRKGGTRALLEMLNHPAVVYKASQEIHFDDNENYKGI 135
Db 46 DVRDGVAPNGSAQQLPQTIIIGVRKGGTRALLEMLSLHPDVAANEVHFFDWEHSHG 105
QY 136 IEWYRKMPFSYPOQITTEKSPAYITEVPRYIKYNSIKLLIIVREPTTRAIISDYTVQ 195
Db 106 LGWYLSQMPFSWPHQITVEKTPAYFTSPKVPERVYSMPNSIRLLILRLDPSERVLSDYTVQ 165
QY 196 VLEGERKRNKTYKPEKLAIDNTECVNTKTKAVRTSIYTKHLEBWLKVFTEQVHVVDG 255
Db 166 VFYNHMQHKPYPSIEEFLVRDG--RLNVVDYKALNRSLYHVHMQNWLRFPPFLRHHIVD 223
QY 256 DRLITEPELQVLEKFLNLPRIISQYNLYFNATRGFYCLRENIFNFKLAGSKGRIHPE 315
Db 224 DRLIDPPEFIOKVERFLKLSQINASNFYFNKTKGYCLR--DSGRDCLHESKGRAHPQ 282
QY 316 VDPSPVITKIKKFHFFPNOKFYQITG 340
Db 283 VDPKLLINKHEYFHEPNKFFELV 307

Db 46 DVDRGVAPNGSAQQLPOTIIIGVRKGGTRALLEMLSLHPDVAAAEVHFDFWEEHSHG 105

QY 136 IEWYRKMPFSYPOQITIEKSPAYFITEEVPRIYKNSIKLIIIVREPTTTRAI SDYTQ 195

Db 106 LGWYLSQMPFSWPHQLTVEKTPAYFTSPKPERVYSNNPSIRULLILRDPSEVLS DYTQ 165

QY 196 VLEGERKNKYIYKFEKLADPNTCEVNTKYKAVRTSIYTKHLERWLKYPPIBQFHV DVG 255

Db 166 VFYNNHMQKHFPYSIEEFLVRDG--RLNVDYKALNRSLYHVHMQNWLRFPLRH IHDVG 223

QY 256 DRLITEPLPELQVKEFLNLPRIISQYNLYFNATRGVYCLRFNIIFNKCLAGSKGR IHP 315

Db 224 DRLIRDFFPEIQVERFLKSPQINASNIFYNKTGFCYCLR-DSGRDCLHESKGR AHPQ 282

QY 316 VDPESVITKLKRFHPFNQKFYQITG 340

Db 283 VDPKLLNKLHGFHEPNKFFELVG 307

RESULT 11

AAM56899

ID AAM56899 standard; Protein; 307 AA.

XX

AC AAM56899;

XX

DT 05-NOV-2001 (first entry)

XX

DE Human brain expressed single exon probe encoded protein SEQ ID NO: 29004.

XX

KN Human; brain expressed exon; gene expression analysis; probe;

KN microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KN epilepsy; cancer.

XX

OS Homo sapiens.

XX

PN WC200157275-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00667.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-483446/52.

XX

PT Single exon nucleic acid probes for analyzing gene expression in human brains -

XX

PS Example 4; SEQ ID NO: 29004; 650pp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.

XX

SQ Sequence 307 AA;

Query Match 41.8%; Score 758; DB 22; Length 307;

Best Local Similarity 46.2%; Pred. No. 7e-67;

Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;

QY 16 LGSVAVGSLIYLVARVGSLSRLQPIPIEGRLGARTQAEFFLRALQFKGLLHPRKGN 75

Db 7 MAALLGAVL-LVAQ----PQLVPSRP-----AELGOELLRKAGTLD----- 45

QY 76 ASKEQVRLHDLVQOLPKAIIIGVRKGGTRALLEMLSLHPDVAAAEVHFDFWEEHSHG 135

Db 46 DVDRGVAPNGSAQQLPOTIIIGVRKGGTRALLEMLSLHPDVAAAEVHFDFWEEHSHG 105

QY 136 IEWYRKMPFSYPOQITIEKSPAYFITEEVPRIYKNSIKLIIIVREPTTTRAI SDYTQ 195

Db 106 LGWYLSQMPFSWPHQLTVEKTPAYFTSPKPERVYSNNPSIRULLILRDPSEVLS DYTQ 165

QY 196 VLEGERKNKYIYKFEKLADPNTCEVNTKYKAVRTSIYTKHLERWLKYPPIBQFHV DVG 255

Db 166 VFYNNHMQKHFPYSIEEFLVRDG--RLNVDYKALNRSLYHVHMQNWLRFPLRH IHDVG 223

QY 256 DRLITEPLPELQVKEFLNLPRIISQYNLYFNATRGVYCLRFNIIFNKCLAGSKGR IHP 315

Db 224 DRLIRDFFPEIQVERFLKSPQINASNIFYNKTGFCYCLR-DSGRDCLHESKGR AHPQ 282

QY 316 VDPESVITKLKRFHPFNQKFYQITG 340

Db 283 VDPKLLNKLHGFHEPNKFFELVG 307

RESULT 12

AAM69287

ID AAM69287 standard; Protein; 307 AA.

XX

AC AAM69287;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29593.

XX

KN Human; bone marrow expressed exon; gene expression analysis; probe;

KN microarray; cancer; leukaemia; lymphoma; myeloma.

XX

OS Homo sapiens.

XX

PN WC200157276-A2.

XX

PD 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00668.

XX

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX

PA (MOLE-) MOLECULAR DYNAMICS INC.

XX

PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX

DR WPI; 2001-488900/53.

XX

PT Human genome-derived single exon nucleic acid probes useful for

PT analyzing gene expression in human bone marrow -

XX

PS Example 4; SEQ ID NO: 29593; 658pp + Sequence Listing; English.

XX

CC The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention.

XX

SQ Sequence 307 AA;
 Query Match 41.8%; Score 758; DB 22; Length 307;
 Best Local Similarity 46.2%; Pred. No. 7e-67;
 Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;

QY 16 LGSIAVGSLLYLVARVGSLLDRLOPICPIEGRGGARTOAEFPLRALQKRGLLHFEFRGN 75
 DB 7 MAALLLGAUL-LVAQ-----POLVPSRP-----AELGQQLLRKAGTLQD----- 45
 QY 76 ASKEQVRLHDLVQOLPKAIIIGVRKGGTRALLEMLNLHPAVVKASQEIHFDDNDENYKGG 135
 DB 46 DVRDGVAPNGSAQQLPQTIIIGVRKGGTRALLEMLSLHPDVAANEVHFDFWEHSHG 105
 QY 136 IEWTRKOMPSPYPOITIEKSPAYFITEEVEPERIYKSNSSIKLLIIVREPTTRAIISDTQ 195
 DB 106 LGWYLSQMPFSPWPHOLTVEKTPAYFTSPKVPERYSMNPSIRLLILLRDPSEVLSDTQ 165
 QY 196 VLEGKERKNTYYKFEKLAIDPNTCEVNTKYKAVRTSYTKHLERWLKYPFIEQFHVVDG 255
 DB 166 VFYNHMQHKYPSTIEEFLVRDG--RLNVDYKALNRSYHVMQNLWLFPLRHHIIVDG 223
 QY 256 DRLITEPELQVKEKFLNLPRIISOYNLYFNATRGFYCLRFNIIFNKCLAGSKGRIHPE 315
 DB 224 DRLIRDPPFPIQKVERFLKLSQINASNIFYNKTGKGYCLR-DGDRCLHESKGRAHPQ 282
 QY 316 VDPSTVITKLKRFHFNQKFYQITG 340
 DB 283 VDPKLLNKLHGFHEPNKGFELVG 307

RESULT 13
 AAM29607
 ID AAM29607 standard; Protein; 307 AA.
 XX
 AC AAM29607;
 DT 17-OCT-2001 (first entry)
 XX
 DE Peptide #3644 encoded by probe for measuring placental gene expression.
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX Homo sapiens.
 OS
 FN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488897/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID No 29876; 654pp; English.
 XX
 CC The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one

CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 SQ Sequence 307 AA;
 Query Match 41.8%; Score 758; DB 22; Length 307;
 Best Local Similarity 46.2%; Pred. No. 7e-67;
 Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;

QY 16 LGSIAVGSLLYLVARVGSLLDRLOPICPIEGRGGARTOAEFPLRALQKRGLLHFEFRGN 75
 DB 7 MAALLLGAUL-LVAQ-----POLVPSRP-----AELGQQLLRKAGTLQD----- 45
 QY 76 ASKEQVRLHDLVQOLPKAIIIGVRKGGTRALLEMLNLHPAVVKASQEIHFDDNDENYKGG 135
 DB 46 DVRDGVAPNGSAQQLPQTIIIGVRKGGTRALLEMLSLHPDVAANEVHFDFWEHSHG 105
 QY 136 IEWTRKOMPSPYPOITIEKSPAYFITEEVEPERIYKSNSSIKLLIIVREPTTRAIISDTQ 195
 DB 106 LGWYLSQMPFSPWPHOLTVEKTPAYFTSPKVPERYSMNPSIRLLILLRDPSEVLSDTQ 165
 QY 196 VLEGKERKNTYYKFEKLAIDPNTCEVNTKYKAVRTSYTKHLERWLKYPFIEQFHVVDG 255
 DB 166 VFYNHMQHKYPSTIEEFLVRDG--RLNVDYKALNRSYHVMQNLWLFPLRHHIIVDG 223
 QY 256 DRLITEPELQVKEKFLNLPRIISOYNLYFNATRGFYCLRFNIIFNKCLAGSKGRIHPE 315
 DB 224 DRLIRDPPFPIQKVERFLKLSQINASNIFYNKTGKGYCLR-DGDRCLHESKGRAHPQ 282
 QY 316 VDPSTVITKLKRFHFNQKFYQITG 340
 DB 283 VDPKLLNKLHGFHEPNKGFELVG 307

RESULT 14
 ABG38901
 ID ABG38901 standard; Peptide; 307 AA.
 XX
 AC ABG38901;
 DT 19-AUG-2002 (first entry)
 XX
 DE Human peptide encoded by genome-derived single exon probe SEQ ID 28566.
 XX
 KW Human; single exon probe; asthma; lung cancer; COPD; ILD;
 KW chronic obstructive pulmonary disease; interstitial lung disease;
 KW familial idiopathic pulmonary fibrosis; neurofibromatosis;
 KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;
 KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;
 KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
 KW primary alveolar proteinosis; fibrocystic pulmonary dysplasia;
 KW hyaline membrane disease.
 XX
 OS Homo sapiens.
 FN WO200186003-A2.
 XX
 PD 15-NOV-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00665.
 XX
 PR 04-FEB-2000; 2000US-180312P.
 PR 26-MAY-2000; 2000US-207456P.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-234687P.
 PR 27-SEP-2000; 2000US-236359P.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WPI; 2002-114183/15.
XX PT Spatially-addressable set of single exon nucleic acid probes, used to
XX PT measure gene expression in human lung samples -
XX PS Claim 27; SEQ ID No 28566; 634bp; English.
XX CC The invention relates to a spatially-addressable set of single exon
XX CC nucleic acid probes for measuring gene expression in a sample derived
XX CC from human lung comprising single exon nucleic acid probes having one of
XX CC 12614 nucleic acid sequences mentioned in the specification, or their
XX CC complements or the 12387 open reading frames derived from the 12614
XX CC probes. Also included are a microarray comprising the novel set of
XX CC probes; the novel set of probes which hybridise at high stringency to a
XX CC nucleic acid expressed in the human lung; measuring gene expression in a
XX CC sample derived from human lung, comprising (a) contacting the array with
XX CC a collection of detectably labeled nucleic acids derived from human lung
XX CC mRNA, and (b) measuring the label detectably bound to each probe of
XX CC the array; identifying exons in a eukaryotic genome, comprising
XX CC (a) algorithmically predicting at least one exon from genomic sequences
XX CC of the eukaryote; and (b) detecting specific hybridisation of detectably
XX CC labeled nucleic acids from eukaryotic lung mRNA, to a single exon probe,
XX CC having a fragment identical to the predicted exon, the probe is included
XX CC in the above mentioned microarray; assigning exons to a single gene,
XX CC comprising (a) identifying exons from genomic sequence by the method
XX CC above and (b) measuring the expression of each of the exons in several
XX CC tissues and/or cell types using hybridisation to a single exon
XX CC microarrays having a probe with the exon, where a common pattern of
XX CC expression of the exons in the tissues and/or cell types indicates that
XX CC the exons should be assigned to a single gene; a peptide comprising one
XX CC of 12011 sequences, mentioned in the specification, or encoded by the
XX CC probes/open reading frames (ORF). The probes are used for gene
XX CC expression analysis, and for identifying exons in a gene, particularly
XX CC using human lung derived mRNA and for the study of lung diseases
XX CC such as asthma, lung cancer, chronic obstructive pulmonary disease
XX CC (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary
XX CC fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease,
XX CC Niemann-Pick disease, Hermansky-Pudlak syndrome, sarcoidosis, pulmonary
XX CC haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis,
XX CC pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic
XX CC pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension
XX CC and hyaline membrane disease. The present sequence is a peptide/protein
XX CC encoded by a single exon probe of the invention.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 307 AA;
Query Match 41.8%; Score 758; DB 23; Length 307;
Best Local Similarity 46.2%; Pred. No. 7e-67;
Matches 150; Conservative 63; Mismatches 88; Indels 24; Gaps 6;
QY 16 LGSVAVGSLYLVARVGSRLDLOPICIEGRGAGARTQAEFFLRALQFRRGLLHFRKGN 75
DB 7 MAALLIGAVL-LVAQ----PQLVPSRP-----AELGQQLLRKAGTLQD----- 45
QY 76 ASKEQVRLHDLVQOLPKALLIGVRKGGTRALLEMLNHPAVKASQEIHFNDENYKGG 135
DB 46 DVRDGVAPNGSAQLOPQTIIIGVRKGGTRALLEMLSLHPDVAARENEVHFFDWEHYSHG 105
QY 136 IEWYRKMPFSPYQQTIEKSPAYFITEEVPRIYKMSIKLLIIVREPTTRTASDYTQ 195
DB 106 LGHYLSQMFSPWQHUTVETKTPAYTSKPKVPIYMSNFSIRULLILRDPSEVSQDYTQ 165
QY 196 VLEGKRNKNTYYKFKLAIDNTCVNTKYKAVRTSIYTKHLERLWKFYFIEQHFVVDG 255
DB 166 VFYNHMQKHKPYPSIEEFLVRDG--RLNVDYKALNESLYHVMQNLRFPLRHHIVDG 223

QY 256 DRLLTEPELPQLVEKFLNLPRIISQYLNFNATRGFYCLRNIIIFNKCLAGSKGRIHPE 315
DB 224 DRLLIRDPPEIQKVERFLKLSQINASFNFKTGFKYCLR-DSGRDRCLHESKGRAPQ 282
QY 316 YDPSVITKLRFFHFPNOKFYQITG 340
DB 283 VDFKLNLKLHFEYHFBPNKKFFELVG 307
RESULT 15
AAV17062
ID AAY17062 standard; Protein; 311 AA.
AC AAY17062;
DT 20-JUL-1999 (first entry)
XX Mouse 3-OST-1 protein.
DE Heparan sulfate D-glucosaminyl 3-O-sulfotransferase; anticoagulant;
KW saccharide; glycosaminoglycan; proteoglycan; antithrombin; 3-OST;
KW heparan sulfate; thrombotic disorder; deep vein thrombosis;
KW pulmonary embolism; coagulation enzyme inactivation.
XX Mus musculus.
OS WO9922005-A2.
PN 06-MAY-1999.
XX 23-OCT-1998; 98WO-US22597.
XX 31-OCT-1997; 97US-0065437.
XX 24-OCT-1997; 97US-0062762.
XX (MASI) MASSACHUSETTS INST TECHNOLOGY.
PI Fritz LMS, Liu J, Rosenberg RD, Schwartz JJ, Shworak NW;
PI Zhang L;
XX WPI: 1999-312968/26.
DR N-PSDB; AAX37245.
XX Heparan sulfate D-glucosaminyl 3-O-sulfotransferase and related
PT polynucleotides
XX Claim 17; Page 71-72; 95pp; English.
XX The invention relates to nucleic acid molecules (AAX37245-X37250)
CC encoding heparan sulfate D-glucosaminyl 3-O-sulfotransferases (3-OSTs).
CC The 3-OST proteins can be used for 3-O-sulfating saccharide residues
CC within a preparation of glycosaminoglycan or proteoglycan
CC polysaccharides. 3-OST-1 can be used for enriching the
CC antithrombin-binding fraction in a preparation of heparan sulfates (HS).
CC 3-OST-1 can also convert HS proteoglycan anticoagulant precursor to the
CC HS anticoagulant (HSACT). The 3-OSTs (Optionally lacking enzymatic
CC function) can be used to determine partial sequence information for
CC complex polysaccharides. The 3-OST proteins, genes and antibodies are
CC also useful for diagnosis of disorders involving HS biosynthesis. 3-OSTs
CC are especially used to generate anticoagulant pentasaccharides, which may
CC be used to treat thrombotic disorders such as deep vein thrombosis and
CC pulmonary embolism. Coagulation enzyme inactivation by antithrombin is
CC enhanced by complexing of antithrombin with endothelial cell surface HS
CC proteoglycans. This is responsible for the non-thrombogenic properties
CC of blood vessels. The present sequence represents a mouse 3-OST-1.
XX Sequence 311 AA;
Query Match 41.8%; Score 757.5; DB 20; Length 311;
Best Local Similarity 48.8%; Pred. No. 8e-67;
Matches 143; Conservative 54; Mismatches 77; Indels 19; Gaps 3;
QY 69 HFRKGNASKQVRLHDLV-----QQLPKALLIGVRKGGTRALLEMLNL 112

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 16:43:02 ; Search time 20 Seconds
(without alignments)
1663.718 Million cell updates/sec

Title: PCT-US03-21094-2
Perfect score: 1814
Sequence: 1 MLFKQAWLRQKLLVLGSLA.....FFHFFNQKFYQITGRTLNWP 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76.*
1: Pirl.*
2: Pirl2.*
3: Pirl3.*
4: Pirl4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	479	26.4	276	2	T33493	hypothetical prote
2	328	18.1	819	2	T29486	hypothetical prote
3	326.5	18.0	883	2	A49733	[heparan sulfate]-
4	316.5	17.4	882	2	A42855	N-heparan sulfate
5	312.5	17.2	882	2	A57169	[heparan sulfate]-
6	172	9.5	308	2	C95935	probable sulfotran
7	143.5	7.9	137	2	A84072	hypothetical prote
8	102	5.6	329	2	H96621	hypothetical prote
9	101.5	5.6	378	2	A90280	hypothetical prote
10	101.5	5.6	383	2	T51466	hypothetical prote
11	99	5.5	803	2	B56277	DNA-directed DNA p
12	99	5.5	853	2	S34682	probable transport
13	98.5	5.4	315	2	H70350	ADP-heptose syntha
14	98.5	5.4	766	2	A86198	hypothetical prote
15	98.5	5.4	841	2	S34624	glycogen phosphory
16	98.5	5.4	1495	2	S61023	hypothetical prote
17	98	5.4	470	2	A90083	hypothetical prote
18	98	5.4	502	2	S64928	hypothetical prote
19	97.5	5.4	692	2	S56849	probable membrane
20	97	5.3	247	1	D24706	modulation protein
21	97	5.3	247	2	H95319	phosphoadenyl-su
22	97	5.3	287	2	JC4531	alcohol sulfotrans
23	97	5.3	430	2	B84946	peptidylprolyl iso
24	96	5.3	1279	2	T41389	rna binding protei
25	94.5	5.2	1361	2	A22959	DNA-directed RNA p
26	94	5.2	796	2	S66769	probable membrane
27	93.5	5.2	442	2	C84985	trigger factor lim
28	93.5	5.2	631	2	G70188	transcription init
29	93.5	5.2	1181	2	B64516	hypothetical prote

30	93.5	5.2	3351	2	T13812	lipophorin - fruit
31	93	5.1	393	2	E97212	glycosyltransferas
32	92.5	5.1	585	2	S77114	ABC transporter sl
33	92.5	5.1	665	1	H97093	fructose-bisphosph
34	92	5.1	458	2	R83912	NADP-specific giut
35	92	5.1	1125	1	F70177	transcription-repa
36	92	5.1	1603	1	VJKW5	vitellogenin vit-5
37	92	5.1	1603	2	F89497	protein vit-5 limp
38	92	5.1	2004	2	AC0314	probable membrane
39	92	5.1	7829	2	T15789	hypothetical prote
40	91.5	5.0	709	2	S03812	exonuclease ABC c
41	91.5	5.0	761	2	G70393	conserved hypother
42	91.5	5.0	842	1	A27335	glycogen phosphory
43	91.5	5.0	1065	2	T37587	probable SMC famil
44	91.5	5.0	1513	2	S45768	mitotic spindle pr
45	91	5.0	339	2	T25315	hypothetical prote

ALIGNMENTS

RESULT 1

T33493
hypothetical protein F40H3.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T33493
E:Gattung, S.
A:Description: The sequence of C. elegans cosmid F40H3.
A:Reference number: Z21358
A:Accession: T33493
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-276 <GAT>
A:Cross-references: EMBL:AF098987; PIDN:AAC67432.1; GSPDB:GN00020; CBSP:F40H3.5
A:Experimental source: strain Bristol N2; clone F40H3
C:Genetics:
A:Gene: CBSP:F40H3.5
A:Map position: 2
A:Introns: 49/2; 109/3; 154/3; 227/3; 271/1

Query Match	26.4%	Score	479;	DB	2;	Length	276;
Best Local Similarity	40.9%	Pred. No.	6.3e-30;				
Matches	106;	Conservative	44;	Mismatches	85;	Indels	24;
Gaps	7;						
Qy	59	RALQFKRGLLHFRKGNASKEQVELHDLVQOLPKAIIIGVRKGGTRALLEMLNHPAVVK	118				
Db	28	RNFKFWI--DLCKGEKS-----HLEKFPSSALIVGVRKGGTRALLDAIALHPKVRI	78				
Qy	119	ASQIHFPDNDENYKGIWYRKOMP-FSPQQOITIEKSPAYFITEEVPRIYKNNSSIK	177				
Db	79	VRRETHFFDS--NTLGFWDYRDQMPVENNDNEIVIEKTPAYFTNEHVPKKVIENPNPMK	136				
Qy	178	LLIIVRPETTRAIADYTVLKGKKNKY-----YKFKLAIDPNTCEVNTKVKAVR	230				
Db	137	LILIVRPVTVSDFTQVYNNKLEQNKLTPVLSVEAFKNEAGIE-----KINMEYKMT	192				
Qy	231	TSIYTKHLERMLKYPFIEQHFVVDGRLITPELQVLEKPLNLPPIISQVNYFNATR	290				
Db	193	NSLYDVHISKVLKFDLKNFLVNGDVFANPLRELKVEEFLGLSRTSITPSQLVFDYNK	252				
Qy	291	GFYCLIRNIFNKLGLASK	309				
Db	253	GFPCFR-KTTKVRCLGLSK	270				

RESULT 2

T29486
hypothetical protein F08B4.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T29486

C:Accession: A57169; G02129; G01581
R:Dixon, J.; Loftus, S.K.; Gladwin, A.J.; Scambler, P.J.; Wasmuth, J.J.; Dixon, M.J.
Genomics 26, 239-244, 1995
A:Title: Cloning of the human heparan sulfate-N-deacetylase/N-sulfotransferase gene from
A:Reference number: A57169; MUID:95324914; PMID:7601448
A:Accession: A57169
A:Molecule type: mRNA
A:Residues: 1-882 <DIX>
A:Cross-references: GB:U18918; NID:9976371; PIDN:AAA75281.1; PID:9976372
A:Note: authors translated the codon AAG for residue 42 as Leu, AAA for residue 110 as I
as Ile
R:Humphries, D.E.
submitted to the EMBL Data Library, September 1995
A:Reference number: G09202
A:Accession: G02129
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-882 <HUM>
A:Cross-references: EMBL:U36600; NID:gl036796; PIDN:AAC27354.1; PID:gl036797
R:Labell, T.L.
submitted to the EMBL Data Library, December 1994
A:Reference number: G07829
A:Accession: G01581
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-59, 'A', 'G', '690-742', 'R', '744-882 <LAB>
A:Cross-references: EMBL:U19790; NID:G841163; PIDN:AAA67765.1; PID:G841164
C:Genetics:
A:Gene: GDB:HSST; NST1
A:Cross-references: GDB:593916; OMIM:600853
A:Map position: 5q32-5q33.1
C:Function: <DAC>
A:Description: as glycosaminoglycan N-acetylglucosaminyl N-deacetylase, hydrolyzes the N-
A:Pathway: heparan sulfate biosynthesis
C:Function: <NST>
A:Description: as desulfoglycan sulfotransferase, catalyzes the formation of glucosamin
enosine-5'-phosphosulfate and converting it to adenosine 3',5'-bisphosphate
A:Pathway: heparan sulfate biosynthesis
C:Superfamily: Caenorhabditis elegans hypothetical protein F08B4.6
C:Keywords: glycoprotein; Golgi apparatus; hydrolase; sulfotransferase; transmembrane pr
F:1-17/Domain: intracellular #status predicted <INT>
F:18-39/Domain: transmembrane #status predicted <TRM>
F:40-882/Domain: trans-Golgi network lumenal #status predicted <LUM>
F:231,351,401,667/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 17.2%; Score 312.5; DB 2; Length 882;
Best Local Similarity 30.6%; Pred. No. 2.8e-16;
Matches 92; Conservative 41; Mismatches 113; Indels 55; Gaps 11;
QY 77 SKEQVRLHDLVQQLPKAIIIGVRKGGTRALLEMLNLHFAV-----VKASQEIHFNDDE 130
DB 596 SKEKT-----CDRFFKLLIIGQKGTGTALYFLGMDPSNPSSTFEEIQFF-NGH 649
QY 131 NYGKIEWYKMPSPYQQIT-----IEKSPAYFITEEVPRIYKNGSIIKLLIIVREPT 186
DB 650 NYHKGIDWNEFFPI--PSNTTSDYFEKSNYFDFSEVAPRAAALLPKAVLTILNPA 707
QY 187 TRAIISDYTVQLGKGRKNTYKFKLAIDPNT-----CEVNTKYKAVRTSI- 233
DB 708 DRAYS-----WYHQRAHDDPVALKYTFHEVITAGSDASSKLRLAQNCL 752
QY 234 ----YTKHLERWLKYPFIEOHVVDGRLITPELQVKEFLNLPRIQY-NLYNA 288
DB 753 VEGWATIERWLSAYHANQILVLDGKLRTPEFAKMDVMQKFLGVTWTDYHKTAFDP 812
QY 289 TRGFYCLRNIIFNKLAGSKGRIHPEVDPSVITKLRKFFHFN--OKFYQITGRTI-N 344
DB 813 KGFVQWQLLEGKTKLGKSKRGKTPEMDLDSRAFLKDYRDHNIELSKLLYKMGQTILPT 872
QY 345 W 345
DB 873 W 873

RESULT 6

C95935
probable sulfotransferase protein [imported] - Sinorhizobium meliloti (strain 1021) mag
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: C95935
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernar
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing end
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: C95935
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-308 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49147.1; PID:g15140632; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler
Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure
hehault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SM521243
A:Genome: plasmid

Query Match 9.5%; Score 172; DB 2; Length 308;
Best Local Similarity 24.5%; Pred. No. 6.3e-06;
Matches 65; Conservative 38; Mismatches 120; Indels 42; Gaps 7;

QY 85 DLVQQLPKA-----IIIGVRKGGTRALLEMLNLHFAVVKASQEIHFNDENYKGI 137
DB 22 EAIEGKPMERDIDFLIIGATKATWLSQLOQDGFMPDPHAFV--SRYYERGD 79
QY 138 WYRKMPSPYQQITIEKSPAYFITEEVPRIYKNGSIIKLLIIVREPTTRAIISDYTV 197
DB 80 WYLSHFAQGEHRLRGKSNYSYMDVPEAAERIEKLPPEARLIAHVNPVDSAYSDYCM 139
QY 198 EGKE--KKNKYTFEKLADPNTCEVNTKYKAVRTSIYTKHLERWLKYPFIEQPHV 255
DB 140 RRAEVGRDIAQY-----LDPRQ---GAGGRFLNGGLYYQQLQYLDLRFPAEQILV 189
QY 256 DRLITEPELQVKEFLNLPRIISQINLYFNATRGFYCLRFNFIKCLAGSKGRIHPE 315
DB 190 EDLXIDARAQLARV-----RGFLGLEADVPLKPLAKKVKDKSEPV 229
QY 316 VDSVITKLRKFFHFNQKPYQITG 340
DB 230 VNPT-LRRLRLPLKPAAPFRQNTG 253

RESULT 7

A84072
hypothetical protein BH3377 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A84072
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A84072
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-137 <STO>
A:Cross-references: GB:AP001518; GB:BA000004; NID:g10175792; PIDN:BA07096.1; GSPDB:GN0
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3377

```
Query Match          7.9%; Score 143.5; DB 2; Length 137;
Best Local Similarity 25.0%; Pred. No. 0.00036;
Matches 38; Conservative 28; Mismatches 59; Indels 27; Gaps 5;

QY 200 KERKNTYKFEKLAIDPNTCEVNTKYKAVRTSTYTKHLERWLKYPPIEQPHVVDGRLI 259
Db 6 KAAKGSGFPSEWV--EKELEGVSGFRFDGRGIYAKQLEQWFRFFPKQFLIKSENFF 63

QY 260 TEPLPELQVKEFLNLP----PRISQYNLYFNATRGFYCLRNIIFNKCLAGSKGRI--H 313
Db 64 EDPKFRKFKYICFLNLPWELPEYVNW-----KVLRSK-RVRY 104

QY 314 PEVDPSVITKLRFFHFNOKFYQITORTLW 345
Db 105 QKINKETRERLLYFKFDFNDQYALINKFGW 136

RESULT 8
H96621
hypothetical protein F23H11.11 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: H96621
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurös, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.B.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: H96621
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-329 <TO>
A;Cross-references: GB:AE005173; NID:g5080813; PIDN:AAD39322.1; GSPDB:GN00141
C;Genetics:
A;Gene: F23H11.11
A;Map position: 1

Query Match          5.6%; Score 102; DB 2; Length 329;
Best Local Similarity 21.8%; Pred. No. 2;
Matches 58; Conservative 38; Mismatches 90; Indels 80; Gaps 13;

QY 130 ENYKGIEWYKMKPSPYQOITIEK-----SPAYITIEVPERIYKMNSSIKLLI 180
Db 30 ESISDLEMDSRM--SRPQIKFEGWSNIQKGITKLIRLEGEPEPTFYFSECFKLYT 87

QY 181 IVREPTTRAIISDYQVLEGEKRNKTYKPEKLAIDNTCEVNTKYKAVRTSTYTKHL- 239
Db 88 IYDMCVQR--SDISQQL-----YEKRVKIED-----YTTQTVLPSEKDED 130

QY 240 -----RWL-KYPPIEQPHVVDGRLITEPLPELQV--EKFLNL-PPRI 279
Db 131 MLRELKVRNNHKTWKLSKFFVYIDRLVRSKI---PIPSIDVGLTCFLDVGDEY 187

QY 280 SQYNLYFNATRGFYCLRNIIF-----NKCLAGSKGRI-----HPEVD 317
Db 188 SYSLSFFNANVFVTHANVLLQVYCEMQSTAKEWIALVEECLMRERVRVTHLSITE 247

QY 318 PSVITKLRFFHFNOKFYQITORTL 343
Db 248 PKLVEKIQ-----NELLVMVTKRL 267

RESULT 9
A90280
hypothetical protein S801258 [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
```

```
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: A90280
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder,
arrett, R.A.; Ragan, M.A.; Senses, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Description: Sulfolobus solfataricus complete genome.
A;Reference number: A99139
A;Accession: A90280
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <KUR>
A;Cross-references: GB:AE006641; NID:g13814452; PIDN:AAK41496.1; GSPDB:GN00155
C;Genetics:
A;Gene: S801258

Query Match          5.6%; Score 101.5; DB 2; Length 378;
Best Local Similarity 22.5%; Pred. No. 2.6;
Matches 81; Conservative 63; Mismatches 137; Indels 79; Gaps 21;

QY 4 KOQAMLRQKLLVLGSLAVGSLLYLV--ARVGLDRLOPTCPIE-----GR 46
Db 36 KDYGW--GEALVAGSLINSYLSINDIAIPLLSRVELSEPIDVTRILEKILFSAGNCGV 93

QY 47 LGARTQAEFPPLAQOFKGLHERFK--GNASKEQVRLHDLVQOLPKA--IIIGVRKGGT 103
Db 94 VSGAISAVEMALWSLKARKSNVELKLGDKIRDSVKYASPFPGKIDVLIATRKSL 153

QY 104 RALLEMLNHP-----AVVKASQEHF-----FDN-----DENYKGI 138
Db 154 RG-FDLVLUHQSPSTVLAIVKAIKREYKVKTAIDNLSFDMLDVAKEFVDFKVKHTEIEW 212

QY 139 YRKQM--PFSYP--QOITTEKSPAYITIEVPERIYKMNSSIKLLIIVREPTTRAI 194
Db 213 IEPLWPNDDYLLSKLT-EFSPPIAAGENESYLGFKKLESGLTYLQPDIAKVG 271

QY 195 QVLEGEKRNKTYKPEKLAID-PNTECVNTYK-----AVRTSYTKHLERNLKYPIEQ 249
Db 272 KELEILDLSK--YKRVLPDRPDASPISTLYTINIGLVKQI--EMVEYTIADFPNDL 327

QY 250 FHVW----DGRLLTEPLPELQVKEFLNLPRIQYNLYFNATRGFYCLRNIIENKCL 305
Db 328 FYSLPKFDG---YISPPENLEVVER-----NIEKIS-YKRLR---ILHFSLE 374

RESULT 10
T51466
hypothetical protein K10A8_100 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C;Accession: T51466
R;Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.; Me
submitted to the Protein Sequence Database, August 2000
A;Reference number: Z25394
A;Accession: T51466
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-383 <SAT>
A;Cross-references: EMBL:AL391151
A;Experimental source: cultivar Columbia; BAC clone K10A8
C;Genetics:
A;Map position: 5
A;Introns: 45/2; 74/1; 94/3; 125/2; 167/2; 197/3; 233/3; 360/1
A;Note: K10A8_100

Query Match          5.6%; Score 101.5; DB 2; Length 383;
Best Local Similarity 21.8%; Pred. No. 2.6;
Matches 83; Conservative 60; Mismatches 133; Indels 105; Gaps 20;

QY 1 MLFKQQAWLKRLVLT----GSLAVGSLLYLVAVGSLDRLOPTCPIEGRLGARTQAEF 56
Db 1 MAAKOMEEIQKKRLLSYPRANAPQSLF-----AGMERYALLBWLFFKLLG--DKSPF 53
```

```
QY 57 PLRALQFKGLLHPRKGNASKSQVELHDLVQ-----QLPKAILIGVRKGGT-RALLEM 109
Db 54 SQONLQGDAGVDE-----ETVRIQYLAETAKFLGTPTVDIEAIQGHGYEDRMEM 105
QY 110 LNLHPAVKASQEIHFDDN-----DENYKGIWY-----RKMPFS-----YPOQITTE 154
Db 106 LRNIVDLVEASL---FSDNQESIDQVAKDQLIDAIARQSLFSECKLFPADVQLQ 162
QY 155 KSPAVITTEVPERIYKMSKLLIIVREPPTTRALSDYTVQVLEGKRNKTYKFEKA 214
Db 163 ---SIYPLPDVSELETKLSEQAKILSNLQO---KVDD-----LAAKHAYNP----- 202
QY 215 IDPNTCEVNTKYKA-----VTSIYTKHLERWLKYPFIEQPHVD---GDRLITE- 261
Db 203 -DEETVESQRLARLESFETARAPNTIYTKIIRWTHMVEVPLQHGFGPAANRLLEY 261
QY 262 ----PLPELQVLEKFLNLPRIISQVNLNFYNAVGRFYCLRFNFIIFNKCLA----- 306
Db 262 NMLLKVPCHLLSVILVSTILSPNLYFSDT-----QIWWQQLGNLXNLRDASHA 314
QY 307 ---GSKGRIHPEDVPSVITKL 324
Db 315 LSIQSSGIVAGE--PSSVTRI 333

RESULT 11
B56277
DNA-directed DNA polymerase (EC 2.7.7.7) II - Pyrodicticum occultum
C:Species: Pyrodicticum occultum
C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 20-Jun-2000
C:Accession: B56277
R:Uemori, T.; Ishino, Y.; Doi, H.; Kato, I.
J. Bacteriol. 177, 2164-2177, 1995
A:Title: The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-like DNA poly
A:Reference number: A56277; MUID:95238290; PMID:7721707
A:Accession: B56277
A:Status: preliminary
A:Accession: B56277
A:Molecule type: DNA
A:Residues: 1-803 <UEM>
A:Cross-references: GB:D38574; NID:9807829; PIDN:BAA07580.1; PID:9807830
C:Superfamily: herpesvirus DNA-directed DNA polymerase
C:Keywords: nucleotidyltransferase

Query Match 5.5%; Score 99; DB 2; Length 803;
Best Local Similarity 18.6%; Pred. No. 11;
Matches 63; Conservative 44; Mismatches 106; Indels 126; Gaps 14;

QY 81 VRLHDLVQQLPK-----AIIIGVRKGGTRALLEMLNLHPAVVKASQEIHFDDNEN 131
Db 295 VDLYDAEMPEIKMKLTLEVAEYGLVMKKSERVIEWRIP-----EYWD-DEK 343
QY 132 YKGIWY-----RKMPFSYPOQITIEKSP-----AY 159
Db 344 KQQLERYALDDVRATYGLAEKMLFPAL-QLSTVTGVLDDQVGMGVGFLEWILMRAAY 402
QY 160 FITEVPERIYKMSKLLIIVREPPTTRALSDYTVQVLEGKRNKTYKFP-----EKLA 215
Db 403 DNNELVFNKVRGSGYKGVVVK--PLKGVHENVVVLDDFSSMYSIMIKYNGVPTIVD 460
QY 216 DPNTC-----EVNTKYKAVRTSIYTKHLERWLKYPFIEQPHVDGDRLITEPLPE 265
Db 461 DSEBPKYGGCVAVEVGHFRFRSPGPKFTVLENLLKL-----FNATRGF----- 500
QY 266 LQVLEKFLNLPRIISQVNLN-----FNATRGF-----YCLRF-----N 298
Db 501 RQVKEKMKFPPDSPEYRLYDRQKALKVLANASYGMGWSHARWYCKEACBAVTANGEN 560
QY 299 IIFN-----KCLAGSKGRIHPEDVPSVITKLRF 327
Db 561 LILTAIEYARKLGLKVIYGDTSLSFVYVDKEKVEKLIKF 599

RESULT 12
```

```
S34682
probable transport protein ALD homolog - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YKL188c; hypothetical protein YKL741; peroxisom
C:Species: Saccharomyces cerevisiae
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 21-Jul-2000
C:Accession: S34682; S38020; S38025; S44318; S39544
R:Wieman, S.; Voss, H.; Schwagaer, C.; Rupp, T.; Stegemann, J.; Zimmermann, J.; Grothue
submitted to the EMBL Data Library, July 1993
A:Description: Sequencing and analysis of 51.5 kilobases on the left arm of chromosome :
A:Reference number: S34679
A:Accession: S34682
A:Molecule type: DNA
A:Residues: 1-853 <WIB>
A:Cross-references: EMBL:X74151; NID:9450365; PID:G395237
R:Wieman, S.; Voss, H.; Schwager, C.; Rupp, T.; Grothues, D.; Sensen, C.; Stegemann, J
submitted to the Protein Sequence Database, March 1994
A:Reference number: S37825
A:Accession: S38020
A:Molecule type: DNA
A:Residues: 1-853 <W12>
A:Cross-references: EMBL:Z28188; NID:9486332; PID:G486333; MIPS:YKL188c
A:Experimental source: strain S288C
R:Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.; Guerreiro, P.;
submitted to the Protein Sequence Database, March 1994
A:Reference number: S38024
A:Accession: S38025
A:Molecule type: DNA
A:Residues: 1-853 <MAI>
A:Cross-references: EMBL:Z28188; NID:9486332; PID:G486333; MIPS:YKL188c
A:Experimental source: strain S288C
R:Bossier, P.; Fernandes, L.; Vilela, C.; Rodrigues-Pousada, C.
Yeast 10, 681-686, 1994
A:Title: The Yeast YKL741 gene situated on the left arm of chromosome XI codes for a ho
A:Reference number: S44318; MUID:95028165; PMID:7941751
A:Accession: S44318
A:Molecule type: DNA
A:Residues: 1-853 <BOS>
A:Cross-references: EMBL:X76133; NID:9438236; PIDN:CAA53736.1; PID:G438237
C:Genetics:
A:Gene: SGD:FXA2
A:Cross-references: SGD:S0001671; MIPS:YKL188c
A:Map position: 11L
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
C:Keywords: ATP; P-loop; peroxisome; purine nucleotide binding; transmembrane protein
F:488-722/Domain: ATP-binding cassette homology <ABC>
F:505-512/Region: nucleotide-binding motif A (P-loop)
F:511/Binding site: ATP/GTP (Lys) #status predicted

Query Match 5.5%; Score 99; DB 2; Length 853;
Best Local Similarity 21.1%; Pred. No. 12;
Matches 58; Conservative 46; Mismatches 125; Indels 46; Gaps 12;

QY 39 PICIEGLGGARTQAEPFLRALQFKGLLHFRKGNAS----KEQVRLHDLVQQLPKAI 94
Db 553 QRPYMGNRSTFRQIIYVPSIEQFKERYHNDYDGLDADLIKILQLDLEDLVT-NMSL 611
QY 95 IIGVR--KGGTRALLEMLNLHPAVVKASQEIHFDDN-DENYKGIWY-----WTRKMP 144
Db 612 LLAQRTSKNDQSLSTEDNQSPCAIKVRDAFSIVRNWSEELTIGVQQLAMARMYHKPK 671
QY 145 PSYQQITIEKSPAYFTEEVPERIYK--MNSSKLLIIVREPPTTRALSDYTVQVLEGKER 202
Db 672 FAVLDECTSAVAP-----EMEQRMYENQAQFGISLSVCHRTSLMHHFNLLKFPDGKGG 725
QY 203 KNTYKFP-----EKLAIDPNTCEVNTKYKAVRTSIYTKHLERWLKYPFIEQPHVDG 256
Db 726 ----YQPPNPKNRKCNEKLELN-----AILDQQVPLWR--KLKDLTAKES 770
QY 257 RLITEPLQLVLEKFLNLPRIISQVNLNFYNAVTRG 291
Db 771 NIIRKSETNMLNLFKIED--FKTSKSNALFNANKG 803
```


Db	315	KEGCRDPVRTNFDAPDKVAIQIANDTHPSLAIPELIRILVDLERLDWDKAMDVTVKTCAY	374
Qy	235	TKH-----LERWLKYPFIEQFHVVDGRLITEPLP-HLQLV-----EKFLN-----LPPR	278
Db	375	TNHTVLPEALEFW---PV-----HLMETLLPRHLQIITYEINQRFNVAFAAFPGD	421
Qy	279	ISOYNLYFNATGFCYCLRENIIFNKLKAGSK-----GRIHPEVDPSVITKLKKEF-----H	329
Db	422	VDRLRMSLVEEG-AVKRINMA-HLCIAGSHAVNGVARIHSEILKKTIFK--DFYELEBFH	477
Qy	330	PFNQKFYOITGR	341
Db	478	KFQNTNGITGR	489

Search completed: December 4, 2003, 16:46:05
Job time : 28 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 16:43:02 ; Search time 18 Seconds
(without alignments)

903.958 Million cell updates/sec

Title: PCT-US03-21094-2

Perfect score: 1814

Sequence: 1 MLFKQQAHLRQKLLVLGSLA.....FFHPPNQKPYQITGTLNWP 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	334.5	18.4	883	1 HSS2 HUMAN	P52849 homo sapien
2	326.5	18.0	883	1 HSS2 MOUSE	P2850 mus musculus
3	316.5	17.4	882	1 HSS1 RAT	Q02353 rattus norv
4	312.5	17.2	882	1 HSS1 HUMAN	P52848 homo sapien
5	104	5.7	247	1 NOH4 RHIME	P06237 rhizobium m
6	99	5.5	853	1 PXA2 YEAST	P34230 saccharomyc
7	98.5	5.4	841	1 PHS2 RAT	P09812 rattus norv
8	97.5	5.4	472	1 PAPI POWPV	Q91588 fowipox vir
9	97.5	5.4	692	1 YH3 YEAST	P40358 saccharomyc
10	97	5.3	247	1 NODH RHIME	P06236 rhizobium m
11	97	5.3	286	1 SUHE CAVPO	P52841 cavia porce
12	97	5.3	430	1 SURA BUCAI	P57240 buchnera ap
13	97	5.3	1048	1 AGO1 ARATH	O04379 arabidopsis
14	95	5.2	803	1 GYRE BUCAP	P29435 buchnera ap
15	94.5	5.2	1361	1 RPOD SPIOL	P11704 spinacia ol
16	94	5.2	796	1 DEC1 YEAST	Q12387 saccharomyc
17	93.5	5.2	442	1 TIG BUCAI	P57546 buchnera ap
18	93.5	5.2	631	1 RPSD BORBU	P52323 borrelia bu
19	93.5	5.2	1181	1 YP02 METJA	O60301 methanococ
20	93	5.1	662	1 UVRB THETN	Q8r8m4 thermoanaer
21	92.5	5.1	585	1 YH70 SINY3	P73627 synchocyst
22	92	5.1	1125	1 MFD BORBU	O51568 borrelia bu
23	92	5.1	1603	1 VIT5 CAEEL	P06125 caenorhabdi
24	92	5.1	2004	1 YP73 YERPE	Q8zdj2 yersinia pe
25	91.5	5.0	761	1 YA88 AQUAE	O67178 aquifex aeo
26	91.5	5.0	841	1 PHS2 HUMAN	P11217 homo sapien
27	91.5	5.0	1065	1 SMCS SCHPO	O13710 schizosacch
28	91.5	5.0	1513	1 STU1 YEAST	P38198 saccharomyc
29	91.5	5.0	1603	1 VIT3 CAEEL	Q9n4j2 caenorhabdi
30	91.5	5.0	1603	1 VIT4 CAEEL	P18947 caenorhabdi
31	91	5.0	881	1 YEET HAEIN	P44288 haemophilus
32	91	5.0	1501	1 CDRA CANAL	O42690 candida alb
33	91	5.0	1597	1 CTRO_MOUSE	P49025 mus musculus

RESULT 1

HSS2_HUMAN

ID HSS2_HUMAN STANDARD; PRT; 883 AA.

AC P52849;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Heparin sulfate N-deacetylase/N-sulfotransferase (EC 2.8.2.-) (N-

HSS1) (N-heparin sulfate sulfotransferase) (Glucosaminyl N-

DE deacetylase/N-sulfotransferase).

GN NDST2 OR HSST2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98264822; PubMed=9601056;

RA Humphries D.E., Lanciotti J., Karlinsky J.B.;

RT "CDNA cloning, genomic organization and chromosomal localization of

human heparan glucosaminyl N-deacetylase/N-sulphotransferase-2.";

RL Biochem. J. 332:303-307(1998).

CC -!- FUNCTION: Catalyzes the N-sulfation and N-deacetylation of

glucosamine of the glucosaminoglycan in heparin sulfate.

CC Plays a role in determining the extent of modification of the

polysaccharide chain.

CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE

(BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; U36601; AAC27120.1; -.

DR EMBL; AF042084; AAB97086.1; -.

DR HSSP; P52848; INST.

DR GENE; HGNC:7681; NDST2.

DR MIM; 603268; -.

DR GO; GO:0016021; C:integral to membrane; TAS.

DR GO; GO:0004393; F:heparin N-deacetylase/N-sulfotransferase ac. . .; TAS.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

KW Transferase; Transmembrane; Glycoprotein; Golgi stack; Signal-anchor.

FT DOMAIN 1 18 CYTOPLASMIC (POTENTIAL)

FT TRANSMEM 19 39 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)

FT DOMAIN 40 883 LUMENAL, CATALYTIC (POTENTIAL).

FT CARBOHYD 233 233 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 400 400 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).

RC TISSUE=Liver;
RX MEDLINE=92348437; PubMed=1379236;
RA Hashimoto Y., Orellana A., Gil G., Hirschberg C.B.;
RT "Molecular cloning and expression of rat liver N-heparan sulfate
sulfotransferase";
RL J. Biol. Chem. 267:15744-15750(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=93248194; PubMed=8483907;
RA Wei Z., Swiedler S.J., Ishihara M., Orellana A., Hirschberg C.B.;
RT "A single protein catalyzes both N-deacetylation and N-sulfation
during the biosynthesis of heparan sulfate";
RL Proc. Natl. Acad. Sci. U.S.A. 90:3885-3889(1993).
CC -!- FUNCTION: Catalyzes the N-sulfation and N-deacetylation of
glucosamine of the glycosaminoglycan in heparan sulfate.
CC Plays a role in determining the extent of modification of the
CC polysaccharide chain.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylyl sulfate + [heparan
sulfate]-glucosamine = adenosine 3',5'-bisphosphate + [heparan
sulfate]-N-sulfoglucosamine.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE
(BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M92042; AAA41701.1; -.
DR HSSP; P52848; INST.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase; Transmembrane; Glycoprotein; Golgi stack; Signal-anchor.
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 18 39 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
(POTENTIAL).
FT FT DOMAIN 40 892 LUMENAL, CATALYTIC (POTENTIAL).
FT ACT SITE 614 614 BY SIMILARITY.
FT DISULFID 818 828
FT CARBOHYD 231 231 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 667 667 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 882 AA; 100784 MW; F3AB9263BF6B4345 CRC64;
Query Match 17.4%; Score 316.5; DB 1; Length 882;
Best Local Similarity 32.2%; Pred. No. 1.1e-18;
Matches 93; Conservative 38; Mismatches 127; Indels 31; Gaps 10;
QY 77 SKEQVRLHDLVQOLPKAIIIGVRKGTTRALLEMLNHPAV-----VKASQIHFFDNDE 130
DB 596 SKEK-----CDRFPKLLIIGPKTGTTALYFLGNHPLDSSNYPSSSTFEIQQF-NGH 649
QY 131 NYGKGIEWYRKMPFSYPOQIT-----IEKSPAYFITEEVPFERYIKMNSIKILLIVREPT 186
DB 650 NYHKGIDTWMEFPPI--PNTTSDFFSEKSNFYFSEVAPRAAALLPKAVLTILINPA 707
QY 187 TRAISDYTVQLGKERKNNYKFKLAIDPNTCEVNTKYKAVRTSI-----YTKHLEKW 241
DB 708 DRAYSWYQORAHDDPVALKYTFHEVITAGP---DASSKLRLQNLKCLPFGVWVATHIERW 764
QY 242 LKVFPEIQPHVVDGLIIEPIPELOLVEKFLNLPRIQY-NLYPNATRGFYCLRNFI 300
DB 765 LSAFHANQLVLDGKLLRTEPAKVMQVDFLGVTSVDYHKTLAFDPKKGFWCOLLEG 824
QY 301 FNKCLAGSKGRJHPDPSVITKIRKFFHPFN---OKFYQITGRTI-NW 345
DB 825 KTKCLGSKGRKYPENDLDSRAFLKDYDYRDHIELSKLLYKNGQILPTW 873

RESULT 4
HSS1 HUMAN
ID HSS1 HUMAN STANDARD; PRT; 882 AA.
AC P52848;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Heparan sulfate N-deacetylase/N-sulfotransferase (EC 2.8.2.8) (N-HSST)
DE (HNSST) (Heparan sulfate)-glucosamine N-sulfotransferase) (N-heparan
DE sulfate sulfotransferase) (Glucosaminyl N-deacetylase/N-
DE sulfotransferase).
GN NDST1 OR HSST1 OR HSST.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
TX TISSUE=Placenta;
RX MEDLINE=95324914; PubMed=7601448;
RA Dixon J., Loftus S.K., Gladwin A.J., Scambler P.J., Wasmuth J.J.,
RA Dixon M.J.;
RT "Cloning of the human heparan sulfate-N-deacetylase/N-sulfotransferase
gene from the Treacher Collins syndrome candidate region at 5q32-
q33.1.";
RL Genomics 26:239-244 (1995).
RN [2]
RP SEQUENCE FROM N.A., AND SUBCELLULAR LOCATION.
TX TISSUE=Umbilical vein endothelial cells;
RX MEDLINE=97361922; PubMed=9230113;
RA Humphries D.E., Sullivan B.M., Aleixo M.D., Stow J.L.;
RT "Localization of human heparan Glucosaminyl N-deacetylase/N-
RT sulfotransferase to the trans-Golgi network";
RL Biochem. J. 325:351-357(1997).
RN [3]
RP SEQUENCE FROM N.A.
TX TISSUE=Connective tissue;
RA Label T.L., Milewicz D.J., Bonadio J., Edelhoff S.,
RA Diesteche C.M., Byers P.H.;
RA Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 579-882.
RX MEDLINE=99214127; PubMed=10196134;
RA Kakuta Y., Sueyoshi T., Negishi M., Pedersen L.C.;
RT "Crystal structure of the sulfotransferase domain of human heparan
RT sulfate N-deacetylase/N-sulfotransferase 1.";
RL J. Biol. Chem. 274:10673-10676(1999).
CC -!- FUNCTION: Catalyzes the N-sulfation and N-deacetylation of
glucosamine of the glycosaminoglycan in heparan sulfate.
CC Plays a role in determining the extent of modification of the
CC polysaccharide chain.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylyl sulfate + [heparan
sulfate]-glucosamine = adenosine 3',5'-bisphosphate + [heparan
sulfate]-N-sulfoglucosamine.
CC -!- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE.
CC -!- TISSUE SPECIFICITY: EXPRESSION IS MOST ABUNDANT IN HEART, LIVER
AND PANCREAS.
CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18918; AAA75281.1; -.
DR EMBL; U36500; AAC27354.1; -.
DR EMBL; U17970; AAA67765.1; -.
DR PIR; A57169; A57169.
DR PDB; 16-SEP-99.

[illegible]

```
Db 54 LLLSRELLERACWRYPHSDKXKTVHVGCKINEPQFQSPSPA-----ELTAWPG 104
QY 176 IKLLIIVREPTRAISDYTVQVLEGRKRNKTYKFEKLAIDP-----NCEVNNYK 227
Db 105 LKVLIVIRNTLESLSFYQA-----RQTRQWLQFKSDSSAPPPVMLPFATCEA----- 154
QY 228 AVRTSYTYKHLB-----RWLKYPIEFQHFVVDGRLITEPLPELQVLEKELNLP 276
Db 155 -----YFKAADDHARVVNAPDSGRILRIEYERLLRDPVPCVATVLDLFGAP 201

RESULT 6
PXA2 YEAST
ID PXA2 YEAST STANDARD; PRT; 853 AA.
AC P34230;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxisomal long-chain fatty acid import protein 1 (Peroxisomal ABC transporter 2).
DE PXA2 OR PAT1 OR YKL188C OR YKL741.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=97030294; PubMed=8876235;
RA Shani N., Valle D.;
RT "A Saccharomyces cerevisiae homolog of the human adrenoleukodystrophy transporter is a heterodimer of two half ATP-binding cassette transporters";
RT Proc. Natl. Acad. Sci. U.S.A. 93:11901-11906 (1996).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=9420264; PubMed=8154185;
RA Wiemann S., Voss H., Schwager C., Rupp T., Stegmann J.,
RA Zimmermann J., Grothues D., Sengen C., Erfle H., Hewitt N.,
RA Banrevi A., Ansoorge W.;
RT "Sequencing and analysis of 51.6 kilobases on the left arm of chromosome XI from Saccharomyces cerevisiae reveals 23 open reading frames including the FAS1 gene.";
RL Yeast 9:1343-1348 (1993).
RN [3]
SEQUENCE FROM N.A.
RX MEDLINE=95028165; PubMed=7941751;
RA Bossier P., Fernandes L., Vilela C., Rodrigues-Pousada C.;
RT "The yeast YKL741 gene situated on the left arm of chromosome XI codes for a homologue of the human ALD protein.";
RL Yeast 10:681-686 (1994).
RN [4]
SEQUENCE FROM N.A.
RA Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA Guerreiro P., Rodrigues-Pousada C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [5]
FUNCTION.
RX MEDLINE=96324389; PubMed=8670886;
RA Hettema E.H., van Roermund C.W.T., Distel B., van den Berg M.,
RA Vilela C., Rodrigues-Pousada C., Wanders R.J.A., Tabak H.F.;
RT "The ABC transporter proteins Pat1 and Pat2 are required for import of long-chain fatty acids into peroxisomes of Saccharomycetes cerevisiae.";
RL EMBO J. 15:3813-3822 (1996).
CC -!- FUNCTION: INVOLVED IN THE IMPORT OF ACTIVATED LONG-CHAIN FATTY ACIDS FROM THE CYTOSOL TO THE PEROXISOMAL MATRIX.
CC -!- SUBUNIT: Forms an heterodimer with PAT2.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MDR SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U93584; AB51597.1; -
DR EMBL; X74151; CA52250.1; -
DR EMBL; X76133; CA53736.1; -
DR EMBL; Z28188; CA82031.1; -
DR PIR; S34682; S34682.
DR SGD; S0001671; PXA2.
DR GO; GO:0005779; C:integral peroxisomal membrane; IDA.
DR GO; GO:0004009; E:ATP-binding cassette (ABC) transporter acti. . . ; IDA.
DR GO; GO:0015908; F:fatty acid transport; IDA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR005283; FA transporter.
DR Pfam; PF00005; ABC tran. 1.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR TIGRFAMs; TIGR00954; 3a01203; 1.
DR PROSITE; PS00211; ABC TRANSPORTER 1; FALSE NEG.
DR PROSITE; PS50893; ABC TRANSPORTER 2; 1.
KW ATP-binding; Glycoprotein; Transmembrane; Transport; Peroxisome.
FT TRANSMEM 122 142
FT TRANSMEM 166 186
FT TRANSMEM 269 289
FT TRANSMEM 364 384
FT NP BIND 505 512
FT NP BIND 512 512
SQ SEQUENCE 853 AA; 97126 MW; BDD37FB2A6A220DB CRC64;

Query Match 5.5%; Score 99; DB 1; Length 853;
Best Local Similarity 21.1%; Fred. No. 2.1;
Matches 58; Conservative 46; Mismatches 125; Indels 46; Gaps 12;

QY 39 PICPIEGRIGGARTOAEPLRALQPKGLLHEFRKGNAS----KQVRLHDLVQOLPKAI 94
Db 553 QRPYMGNRSTFREQIIPDSIEQFKRYHNDYDGLADLIKILQLDLELVTE-NMSL 611
QY 95 IIGVR--XGTRALLEMLNHPAVVKASQEIHFDPN-DENYKGKIE-----WYRKMP 144
Db 612 LLAQRTSKNDSQLSTEDNQSPCAIKVRDAFSIVENWSELTIGVQQRLLAMARMYHKPK 671
QY 145 PSYQQTIERSPAYFITEVEPERIYK--MNSSIKLLIIVREPTRAISDYTVQVLEGRK 202
Db 672 FAVLDECTSAVAP-----EMEQRMYENAQNGISLISVCHRTSLWHFHNLLKFDGKG 725
QY 203 KNTKYK-----EKLALDPNTECVNTKYKAVRTSIVTKHLERWLKYFPIEQHFVVDG 256
Db 726 -----YQGFPPNPKERLCNEEKLLELN-----AILDQVPLWER--KLKDLTIKES 770
QY 257 RLITEPLELQVLEKFLNLPPIRISQNYLYFNATRG 291
Db 771 NIIRKSETNLNLFEXIED--PKTSKSNALFNANKG 803

RESULT 7
PXS2 RAT
ID PXS2 RAT STANDARD; PRT; 841 AA.
AC P09812;
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glycogen phosphorylase, muscle form (EC 2.4.1.1) (Myophosphorylase).
GN PYGM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93320110; PubMed=7916624;
RA Hudson J.W., Hefferon K.L., Creer M.M.;
```



```
QY 122 EHFEDNDENYG-----KGIEWRKMPF----- 145
Db 182 SLHLNKNIEGYDVLQTNARTFLINAFILKFTIGRIVLLKVPFLKNVIMHDEETN 241
QY 146 SYPOQITTEKSPAFITEVEPERIYKMSIKLLIIVREPTRAISDYTVQLEGKERKN 205
Db 242 HVMDFNIREKTMWIPKIMIDNMWIVDPCQLNNIK-----MISQIDRLLEE 289
QY 206 TYKPEKLAIDPNTCEVNTKY-----AVRTSI-----YTKHLERWLK- 243
Db 290 LQAKFEKLSVRLGTLLEVTYRYSIPDSILEVRKLDKDKRKITVDFKKYKLYNIKC 349
QY 244 YPPIEQHFVVDGRLITLPELQVKEFLNLPRIISOYNLYFNATRGFYCLR-----FNI 299
Db 350 YFYLDVEV-----LKKFISKNSGLDEYEDFEAVTNSYAIRNKMTWYTY 392
QY 300 IFNKCLAGSKRIHPEVDPSVITLKRKFPHFPNQKFY 336
Db 393 PSNTALMRSENEIHP-ITINALTSHALLYHVITRAFY 428

RESULT 9
YJH3 YEAST
ID YJH3 YEAST STANDARD; PRT; 692 AA.
AC P40358;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical 80.4 kDa protein in SMC3-MRPL8 intergenic region.
GN YJL073W OR J1083 OR HRC558
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Rose M., Koetter P., Entian K.D.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sor F.J.;
RL STRAIN=S288c;
RN [3]
RP SEQUENCE OF 135-692 FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=95282514; PubMed=7762302;
RA Vandenbol M., Durand P., Dion C., Portetelle D., Hilger F.;
RT "Sequence of a 17.1 kb DNA fragment from chromosome X of
RT Saccharomyces cerevisiae includes the mitochondrial ribosomal protein
RT L8."
RL Yeast 11:57-60(1995).
CC -!- SIMILARITY: Contains 1 J domain.
CC -!- SIMILARITY: TO C.ELRGANS F22B7.5.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49348; CAA89365.1; -
DR EMBL; X88851; CAA61312.1; -
DR EMBL; Z34288; CAA84049.1; -
DR PIR; S56849; S56849.
DR HSP; P08622; 1B02.
DR SGD; S0003609; JEM1.
DR GO; GO:0005783; C:endoplasmic reticulum; IDA.
DR GO; GO:0000300; C:peripheral membrane protein of membrane fra. . .; IDA.
DR GO; GO:0003767; F:co-chaperone activity; IGI.
```

```
DR GO; GO:0000742; P:karyogamy during conjugation with cellular . . .; IGI.
DR GO; GO:0006457; P:protein folding; IGI.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; FALSE_NEG.
DR PROSITE; PS00076; DnaJ_2; 1.
KW Hypothetical protein.
FT DOMAIN 560 569 POLY-GLN.
FT DOMAIN 585 655 J-DOMAIN.
SQ SEQUENCE 692 AA; 80381 MW; 9F612DD16B66981B CRC64;

Query Match 5.4%; Score 97.5; DB 1; Length 692;
Best Local Similarity 20.2%; Pred. No. 2.1; Indels 107; Gaps 15;
Matches 62; Conservative 48; Mismatches 107; Gaps 15;

QY 68 LHEFRKNGKASKEQVRLHDLVQQLPKAIIIGVRKGGTRALLEMLNLHPAVVKASQBIHFDD 127
Db 313 ISKLKNVPSKQI-----LDPAT-----YAFE 335
QY 128 NDE--NYGKGIWYRK-KMPSPYQOI-----TIKSPAYFTTEVEPERIYKMSIKLLI 180
Db 336 NKFRSWDRITIEFYLDKDKKPFITPMKILNKDTNFKNNYFFLEEIKQLIEDVQLS----- 390
QY 181 IVREPTTALSDYTVQLEGKERKNKYKPEKLA-IDPNTCEVNTK----- 225
Db 391 ---RPLAKNLFEDPPITDGFV-KPKSYHTDLYVIDLSILCOASSMSPDVKRAKLAAPFC 446
QY 226 YKAVRTSIYTKHLERWLKYPFIEQHFVVDGRLITLPELQVKEFLNLPRIISOYNLY 285
Db 447 KCSLRHSL---TLETWKHY-----QDAKSKQPLPEITVLSDVW-NSNPHLLMYWV- 492
QY 286 FNATRGFYCLRNFIFNKCLAGSKRIHPEVDPSVITLKRKF-----HFPNQKFY 336
Db 493 -----NSILNK-----SRSKPESQPKQLYDQINKPFQDNGLSESTNPTVYMKNF 536
QY 337 QITGRTL 343
Db 537 RLLOKQL 543
```

RESULT 10

```
NODH RHIME
ID NODH RHIME STANDARD; PRT; 247 AA.
AC P06236;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nodulation protein H (RC 2.8.2.-) (Host-specificity of nodulation
DE protein D).
GN NODH OR HSND OR RA0464 OR SMA0851.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=87016382; PubMed=3020515;
RA DeBelle F., Sharma S.B.;
RT "Nucleotide sequence of Rhizobium meliloti RCR2011 genes involved in
RT host specificity of nodulation."
RL Nucleic Acids Res. 14:7453-7472(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RA Fisher R.F., Swanson J.A., Mulligan J.T., Long S.R.;
RT "Extended region of nodulation genes in Rhizobium meliloti 1021. II.
RT Nucleotide sequence, transcription start sites and protein products."
RL Genetics 117:191-201(1987).
RN [3]
RP SEQUENCE FROM N.A.
```



```

RX STRAIN=1021;
RX MEDLINE=21396509; PubMed=11481432;
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barlett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barlow-Hubler F., Bowser L., Capella D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kalman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yen K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
CC -!- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.
CC PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO A N-
CC ACETYLGLUCOSAMINE OF THE NOD FACTOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X04380; CAA27963.1; -
CC EMBL; M37417; AAA26339.1; -
CC EMBL; AB007237; AAK65122.1; -
CC PIR; D24706; D24706.
CC PIR; H95319; H95319.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransferase; 1.
KW Nodulation; Transferase; Plasmid; Complete proteome.
FT DOMAIN
FT 1
FT HYDROPHOBIC.
SQ SEQUENCE 247 AA; 28584 MW; 545BC0517BAFEBD3 CRC64;

Query Match 5.3%; Score 97; DB 1; Length 247;
Best Local Similarity 18.0%; Pred. No. 0.57;
Matches 42; Conservative 48; Mismatches 69; Indels 74; Gaps 10;

QY 84 HDLVQOLPKAIIIGVKGKTRALLEMLNHPAVVKASQBIHFDDNDENYKGIWYRK-- 141
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 3 HSTLPQPPA-ILAMPRTGTHYLELVNEHPNLSNGELLNYDTN-----WPKRER 53
QY 142 -----KMPFSYPOQIT-----IEKSPAYFITEVPERIYKNS--S 175
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 54 LLLSDRELERAFLLRYPHSDKKTHTVHGCKINEPQQRPSFFA-----ELTAWPG 104
QY 176 IKLLIVREPTTRALSDYQVLEGEKRNKTYVKEKLAIDP-----NTCEVNTKYK 227
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 105 LKVLIVIRNTIESLSFVQA-----RQTRWLKFKSDSAPPVWMLFPAICEA----- 154
QY 228 AVRTSIYTKHLE-----RWLKYFPIEQFHVVDGRLITEPLPELQVKEFLNLP 276
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 155 -----YFKAADDFHARVVYAFDSSRIILTEVERLLRDPVCPATVLDLFLGAP 201

RESULT 11
SUB_CAVPO STANDARD; PRT; 286 AA.
AC P52841;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 3-beta-hydroxysteroid sulfotransferase (EC 2.8.2.2) (Alcohol
DE sulfotransferase) (Pregnenolone sulfotransferase) (Hydroxysteroid
DE sulfotransferase 2) (HST2).
GN STD2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Adrenal gland;
RX MEDLINE=96268476; PubMed=8672244;

RA Dufort I., Tremblay Y., Belanger A., Labrie F., Luu-The V.;
RT "Isolation and characterization of a stereospecific
RT 3beta-hydroxysteroid sulfotransferase (pregnenolone sulfotransferase)
RT cDNA."
RL DNA Cell Biol. 15:481-487(1996).
RN [2]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NIH 2; TISSUE=Adrenal gland;
RX MEDLINE=96125350; PubMed=8554560;
RA Iau N.X., Driscoll W.J., Martin B.M., Strott C.A.;
RT "Molecular cloning and expression of a guinea pig 3-hydroxysteroid
RT sulfotransferase distinct from chiral-specific 3 alpha-hydroxysteroid
RT sulfotransferase."
RL Biochem. Biophys. Res. Commun. 217:1078-1086(1995).
CC -!- FUNCTION: Catalyzes the sulfation of 3-beta-hydroxyl groups of
CC neutral steroids. High preference for C21 steroid (pregnenolone).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol =
CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: LIVER, INTESTINE AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U55944; AAB07868.1; -
CC EMBL; U35115; AAC52347.1; -
CC PIR; JC4531; JC4531.
CC HSP; P50224; 1CJM.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransferase; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT INIT MET
FT 0
FT BY SIMILARITY.
FT BINDING
FT 248 254 PAPS BINDING SITE (POTENTIAL).
FT CONFLICT
FT 204 204 R -> K (IN REF. 2).
SQ SEQUENCE 286 AA; 33932 MW; 23B8366016CF5A4B CRC64;

Query Match 5.3%; Score 97; DB 1; Length 286;
Best Local Similarity 21.0%; Pred. No. 0.69;
Matches 65; Conservative 42; Mismatches 114; Indels 88; Gaps 13;

QY 56 PPLRALQFKGLLHFFRKGNASKQVRLHDLVQQLPKAIIIGVKGKTRALLEMLNHL 113
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 12 FPM--VGSPPELLREVDRKFLVKDE-----DTITVTPKSGTNWLIIEIVCLILS 58
QY 114 PAVKASQEIHFDDN-----DENYKGIWYKRNKTYVKEKLAIDPNTCEVNTKYK 169
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 59 KGDPKWQSVPTWDRSPWIEQHGNEI-MKSDQPRIYTHSLPLHLPFKSPF----- 108
QY 170 YKMSISKLLIIVREPTTRALSDYQVLEGEKRNKTYVKEKLAIDPNTCEVNTKYK 228
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 109 --FSSKAVIYICINPRDLVSGY-----YFTSKMIAEKFPELQVWKWFL 153
QY 229 VRTSIY---TKHLERWLKYPFIEQFHVVDGRLITEPLPELQVKEFL--NLPP----- 277
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 154 QGVNIYGSWFHVRDMLSMREKNFVLVSVELIKDTRISIVEKIKQGLKGLKPEEIDL 213
QY 278 -----RISQYNLYEN--ATRGFYCLR-----FNIFNKL 305
DQ : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DQ 214 LKYSKFKPMKENEMSNVSLLENLTTGFTFLRKGVVDGDKHFTVAQAEFDFKIYQSKM 273
QY 306 AGSKGRTHP 314
DQ : : : : :
DQ 274 AGYPPKUPP 282
```



```
FT DOMAIN 676 997 PIWI.
FT DOMAIN 13 104 GLY-RICH.
SQ SEQUENCE 1048 AA; 116190 MW; 3E5146343A09C541 CRC64;

Query Match
Best Local Similarity 5.3%; Score 97; DB 1; Length 1048;
Matches 63; Conservative 50; Mismatches 97; Indels 72; Gaps 13;

QY 10 ROKLVGLSVGLVLLVAVVGSIDRLQPICFIEGRIGGARTQAEFFPL-----RALQF 63
D5 251 KSLYTAGELPNSKEF---RINLIDE-----EVGAGQRREREFKVKIVLVARADLH 300
QY 64 KRGLLHFRKGNASKEQVRLHDLV-QQLPKAIIIGVRKGGTTRALLEMLNLHPVAVKASQE 122
D5 301 HLGMELEGKQSDAPQEAQVLDIVLRLEPSTSYIPVGRS-----339
QY 123 IHFFDND-----ENYKGLIE---WYRKMPFYPQOITIEKSPAYFITEEVEPERIYKNS 174
D5 340 --FYSFDIGKKGSGDGLGSEWRFQYSIRPTQMGSLNIDMSSTAFI-----EANP 388
QY 175 SIKLL--IIVRPTTTRAIASDYTVQVLEGRKKNKTYKFEKLAIDNCEVNTKYK-AVRT 231
D5 389 VIQFVCDLLNRDISRPLSDARV-----KIKKALRGVKVAV-THRGWNRKRYISGLT 441
QY 232 STYTKHL-----ERMLKYPPIEQHFVVDGDRILTEPLPELQ 268
D5 442 AVATRELTPPVDERNTQKSVWEYFETHYGRFRIQHTQLPCLQV 483

RESULT 14
GRXB_EUCAP
ID GYRB_EUCAP STANDARD; PRT; 803 AA.
AC P29435;
DT 01-APR-1993 (Rel. 25, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA gyrase subunit B (EC 5.99.1.3).
GN GYRB OR BUSG010.
OS Buchnera aphidicola (subsp. Schizaphis graminum).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Buchnera.
OX NCBI_TaxID=98794;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98184963; PubMed=9516544;
RA Clark M.A., Baumann L., Baumann P.;
RT "Sequence analysis of a 34.7-kb DNA segment from the genome of
RT Buchnera aphidicola (endosymbiont of aphids) containing groEL, dnaA,
RT the atp operon, gldA, and rho.";
RL Curr. Microbiol. 36:158-163 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22084549; PubMed=12089438;
RA Tanas I., Klasson L., Canbaeck B., Naeslund A.K., Eriksson A.-S.,
RA Wernegreen J.J., Sandstrom J.P., Moran N.A., Andersson S.G.B.;
RT "50 million years of genomic stasis in endosymbiotic bacteria.";
RL Science 296:2376-2379 (2002).
RN [3]
RP SEQUENCE OF 1-91 FROM N.A.
RX MEDLINE=92241666; PubMed=1572539;
RA Lai C.-Y., Baumann P.;
RT "Genetic analysis of an aphid endosymbiont DNA fragment homologous to
RT the rnpA-rpmH-dnaA-dnaN-gyrB region of eubacteria.";
RL Gene 113:175-181 (1992).
CC -!- FUNCTION: DNA GYRASE NEGATIVELY SUPERCOILS CLOSED CIRCULAR DOUBLE-
CC STRANDED DNA IN AN ATP-DEPENDENT MANNER AND ALSO CATALYZES THE
CC INTERCONVERSION OF OTHER TOPOLOGICAL ISOMERS OF DOUBLE-STRANDED
CC DNA RINGS, INCLUDING CATENANES AND KNOTTED RINGS.
CC -!- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
CC of double-stranded DNA.
CC -!- SUBUNIT: MADE UP OF TWO CHAINS. THE A CHAIN IS RESPONSIBLE FOR DNA
CC BREAKAGE AND REJOINING; THE B CHAIN CATALYZES ATP HYDROLYSIS. THE
CC ENZYME FORMS AN A2B2 TETRAMER.
```

```
CC -!- SIMILARITY: Belongs to the type II topoisomerase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF008210; AAC38108.1; --
CC EMBL; AE014076; AAM67582.1; --
CC EMBL; M80817; AAR73151.1; --
CC HSSP; P06982; 1AJ6.
CC
CC InterPro; IPR003594; ATPbind_ATPase.
CC InterPro; IPR003288; DNA_gyraseB_C.
CC InterPro; IPR000565; DNA_gyrB.
CC InterPro; IPR001241; DNA_topoisomII.
CC InterPro; IPR006171; Toprim_dom.
CC Pfam; PF00204; DNA_gyraseB; 1.
CC Pfam; PF00986; DNA_gyraseB; 1.
CC Pfam; PF02518; HATPase_C; 1.
CC Pfam; PF01751; Toprim; 1.
CC PRINTS; PR00418; TPI2FAMILY.
CC ProDom; PD149633; DNA_gyraseB_C; 1.
CC SMART; SM00387; HATPase_C; 1.
CC SMART; SM00433; TOP2c; 1.
CC TIGRfam; TIGR01059; gyrB; 1.
CC PROSITE; PS00177; TOPOISOMERASE_II; 1.
KW Topoisomerase; Isomerase; ATP-binding; Complete proteome.
FT CONFLICT 146 146 K -> E (IN REF. 1).
SQ SEQUENCE 803 AA; 92677 MW; P4E2BD97EA93A71 CRC64;

Query Match
Best Local Similarity 5.2%; Score 95; DB 1; Length 803;
Matches 73; Conservative 63; Mismatches 139; Indels 88; Gaps 15;

QY 9 LRQKLVGLSVGLVLLVAVVGSIDRLQPICFIEGR-----46
D5 417 LSEIYLVGDSAGS-----AKQGNKKNQAILPLKGLNLNVQTKPKIILSQELTSLI 471
QY 47 --LGAQTQAEFFPLRALQFKGLLHFRKGNASKEQVRLHDLV-----QQLPKAIIIG-- 97
D5 472 TALGCSITKSEYSLDKLYN---HIIIMTDVGDGAHIRTLLTFFYRQLPELIKGVV 527
QY 98 -----VRKGGTTRALL---EMNLHPAVVAVKASQEHFFPDNDENYKGIENYKMPF 145
D5 528 YIAOPPLYKVKKGQKQKQKYNKQKNDENKQYQIKALKEIVKKNKNCNKIKPKQIVSD 587
QY 146 SYPOQITIEKSPAYFITEEVEPERIYKNSIKLIIIVREPTRAIASDYTVQVLE---GKER 202
D5 588 FNHIQIKMKKNKYF-----PELI--LNELIYHPRLYNLKNENVVQNWIEKLVKLNK 640
QY 203 KKKTYKFEKLAIDPNTCEVNTKYKAVRTSYTGHLEWLK--YFPDQHFVVDGDRILIT 260
D5 641 KNNIYIT-SKIKRND---SIFPSIKLSRYANHQTQYDLKNDFLESKEYFLITN---LG 693
QY 261 EPLPELQVLEKFLNPLPRISQY-----LYFNATRGFYCLFRNIIFNCLAGSKGRI 312
D5 694 EKFKFQENENFIEKGEKIYKINDIKNTLEWLIKETKRGFFVQRYK-----GLGM 744
QY 313 HPE 315
D5 745 NPD 747

RESULT 15
ID _RPD SPIOL STANDARD; PRT; 1361 AA.
AC P11704;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 16:43:02 ; Search time 41 Seconds
(without alignments)
2177.712 Million cell updates/sec

Title: PCT-US03-21094-2
Perfect score: 1814
Sequence: 1 MLFKQQAHLRQKLLVLGSLA.....FFHFNQKFYQITGRTLNWP 346

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES					
Result No.	Score	Query Match	Length DB ID	Description	
1	1814	100.0	346 4	Q81ZT8	Q81ZT8 homo sapien
2	1768	97.5	346 11	Q8BSL4	Q8BSL4 mus musculus
3	935	51.5	175 4	Q8N285	Q8N285 homo sapien
4	780	43.0	307 4	O14792	O14792 homo sapien
5	757.5	41.8	311 11	O35310	O35310 mus musculus
6	745.5	41.1	311 11	Q9ESG5	Q9ESG5 rattus norv
7	726.5	40.0	605 5	Q8MRE7	Q8MRE7 drosophila
8	646.5	35.6	384 5	Q9VWJ7	Q9VWJ7 drosophila
9	643.5	35.5	335 4	Q8NDC2	Q8NDC2 homo sapien
10	637.5	35.1	406 4	Q9Y663	Q9Y663 homo sapien
11	636.5	35.1	390 4	Q9Y662	Q9Y662 homo sapien
12	635.5	35.0	367 4	Q9Y278	Q9Y278 homo sapien
13	615.5	33.9	393 11	Q8BKNE	Q8BKNE mus musculus
14	614.5	33.9	390 11	Q9QZ56	Q9QZ56 mus musculus
15	608	33.5	311 4	Q96RX7	Q96RX7 homo sapien
16	601.5	33.2	250 4	Q9Y661	Q9Y661 homo sapien

17	600	33.1	381	4	Q96QI5	Q96QI5 homo sapien
18	479	26.4	276	5	Q9TZK5	Q9TZK5 caenorhabdi
19	397.5	21.9	172	6	Q9XST1	Q9XST1 canis famil
20	347.5	19.2	872	4	Q9H3R1	Q9H3R1 homo sapien
21	344.5	19.0	873	11	Q9EOH7	Q9EOH7 mus musculu
22	343.5	18.9	872	11	Q9EOW8	Q9EOW8 mus musculu
23	342.5	18.9	458	11	Q9D557	Q9D557 mus musculu
24	341.5	18.8	873	4	Q9S803	Q9S803 homo sapien
25	341.5	18.8	876	4	Q9UE21	Q9UE21 homo sapien
26	339.5	18.7	884	6	Q97583	Q97583 bos taurus
27	330.5	18.2	1048	5	Q9V3L1	Q9V3L1 drosophila
28	328	18.1	696	5	Q966W4	Q966W4 caenorhabdi
29	328	18.1	715	5	Q19197	Q19197 caenorhabdi
30	328	18.1	814	5	Q966W5	Q966W5 caenorhabdi
31	328	18.1	826	5	Q966W6	Q966W6 caenorhabdi
32	328	18.1	852	5	Q966W3	Q966W3 caenorhabdi
33	319.5	17.6	882	11	Q9R206	Q9R206 mus musculu
34	318.5	17.6	882	11	O70353	O70353 mus musculu
35	314.5	17.3	153	11	Q8C055	Q8C055 mus musculu
36	282	15.5	240	5	Q9V8M0	Q9V8M0 drosophila
37	173.5	9.6	247	2	Q44300	Q44300 aeromonas s
38	172	9.5	308	16	Q92VF8	Q92VF8 rhizobium m
39	160	8.8	176	11	O8BLP1	O8BLP1 mus musculu
40	143.5	7.9	137	16	O9K7I4	O9K7I4 bacillus ha
41	139.5	7.7	363	11	O8BLQ5	O8BLQ5 mus musculu
42	139.5	7.7	561	11	Q9D2N6	Q9D2N6 mus musculu
43	139.5	7.7	561	11	O9IXQ5	O9IXQ5 mus musculu
44	138.5	7.6	561	11	Q8CHI9	Q8CHI9 rattus norv
45	131.5	7.2	561	4	O60338	O60338 homo sapien

ALIGNMENTS

RESULT 1

Q81ZT8 PRELIMINARY; PRT; 346 AA.

AC Q81ZT8; 2003 (TREMREL. 23, Created)

DT 01-MAR-2003 (TREMREL. 23, Last sequence update)

DT 01-MAR-2003 (TREMREL. 23, Last annotation update)

DE Heparan sulfate 3-O-S-5.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

RN NCBI_TaxID=9606;

RX MEDLINE=22241877; PubMed=12138164;

RA Xia G., Chen J., Tiwari V., Ju W., Li J.P., Malmstrom A., Shukla D.,

RA Liu J.;

RT "Heparan Sulfate 3-O-Sulfotransferase Isoform 5 Generates Both an

RT Antithrombin-binding Site and an Entry Receptor for Herpes Simplex

RT Virus, Type 1."

RL J. Biol. Chem. 277:37912-37919(2002).

DR EMBL; AF503292; AA037737.1; -.

SQ SEQUENCE 346 AA; 40408 MW; C763F70793FDB156 CRC64;

Query Match 100.0%; Score 1814; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 8.5e-137; Indels 0; Gaps 0;
Matches 346; Conservative 0; Mismatches 0;

Qy	1	MLFKQQAHLRQKLLVLGSLAVGSLILYVARVGSILDRLOPICPIEGRIGGARTQAEFFLRA	60
Db	1	MLFKQQAHLRQKLLVLGSLAVGSLILYVARVGSILDRLOPICPIEGRIGGARTQAEFFLRA	60
Qy	61	LQFKRGLLHFRKGNASKQVRHLVDLVQQLPKAIIIGVRKGGTRALLEMLNHPAVVKAS	120
Db	61	LQFKRGLLHFRKGNASKQVRHLVDLVQQLPKAIIIGVRKGGTRALLEMLNHPAVVKAS	120
Qy	121	QEHFFPDNDENYKGIWYRKQMPFSPQITTEKSPAYFITEVPERIYKNSIKLLI	180
Db	121	QEHFFPDNDENYKGIWYRKQMPFSPQITTEKSPAYFITEVPERIYKNSIKLLI	180

Query Match 43.0%; Score 780; DB 4; Length 307;
Best Local Similarity 46.4%; Pred. No. 2.7e-54;
Matches 153; Conservative 64; Mismatches 89; Indels 24; Gaps 6;
QY 16 LGSIAVGSLLVLVARVGSGLDRLOPICTEIGRGGARTOAEPLRALQPKGLLHEFRKGN 75
D 1 MAALLGAVL-LVAQ----PQLVSRP-----AELGQOELLKAKATLQD----- 39
QY 76 ASKSOVRLHDVQOLPKAIIIGVRKGGTRALLEMLNHPAVVKASOEIHFFDNDENYKGG 135
D 40 DVRCVAPNGSAQQLPOTIIIGVRKGGTRALLEMLSLHPDVAANEVHFFDWEHSHG 99
QY 136 IEWTRKMPSPYPOQITIEKSPAYITEEVEERYKANSKILLIIVREPTTRAISSDTQ 195
D 100 LGWYLSQMPFSPHQLTVEKTPAYTPSPKVPERYVSMNPISRLILLIIRDPSEVLSYDQ 159
QY 196 VLEGKERKNTYYKFEKLAIDPNTCEVNTKVKAVRTSYTKHLERWLKYPFIEQFHVVDG 255
D 160 VFYVNMQKHKYPPISEELVRDG--RLNVDYKALNRSLYHVHMQNWLEPFLRHHIVDG 217
QY 256 DRLITEPELQVKEFLNLPRISSQYNLYFNATRGFYCLRFNIFNKCLAGSKGRHPE 315
D 218 DRLIRDPPEIETQKVERFLKSPQINASNFYFNKTKGFYCLR-DSGRDRCLHESKRAHPQ 276
QY 316 VDPSPVTKLRFKFFHFNQKFIQITGRITLW 345
D 277 VDPKLLNKLHIEFHEPNKFFELVGRITPDW 306

RESULT 5
O35310 PRELIMINARY; PRT; 311 AA.
ID O35310
AC O35310
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Heparan sulfate D-glucosaminyl 3-O-sulfotransferase-1 precursor
DE (Heparan sulfate (Glucosamine) 3-O-sulfotransferase 1).
GN Hs3st1 OR 308r1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C3H/An;
RX MEDLINE=98010647; PubMed=9346953;
RA Shworak N.W., Liu J., Fritze L.M.S., Schwartz J.J., Zhang L.,
RA Leggett D., Rosenberg R.D.;
RT "Molecular cloning and expression of mouse and human cDNAs encoding
RT heparan sulfate D-glucosaminyl 3-O-sulfotransferase.";
RL J. Biol. Chem. 272:28008-28019(1997).
RN [2]
RN SEQUENCE FROM N.A.
RP TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (JUN-2001) to the ENBL/GenBank/DBJ databases.
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Ovary;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
DR EMBL; AF019385; AAB84387.1; -;
DR EMBL; BC009133; AAH09133.1; -;
DR EMBL; AK087753; BAC39991.1; -;
DR HSSP; P52848; INST.
DR MGD; MGI:1201606; Hs3st1.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.

KW Signal; Transferase. 20
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 311 HEPARAN SULFATE D-GLUCOSAMINYL 3-O-
FT Sulfotransferase-1.
SQ SEQUENCE 311 AA; 35899 MW; 7DCAF494C838FD7 CRC64;
Query Match 41.8%; Score 757.5; DB 11; Length 311;
Best Local Similarity 48.8%; Pred. No. 1.7e-52;
Matches 143; Conservative 54; Mismatches 77; Indels 19; Gaps 3;
QY 69 HEFRKGNASKEQVRLHDLV-----QOLPKAIIIGVRKGGTRALLEMLN 112
D 21 HPAAPGGLKQGLLRKVIILPDTGEGTAGNSTQOLPOTIIIGVRKGGTRALLEML 80
QY 113 HPAVVKASOEIHFFDNDENYKGGIWKYKMPSPYPOQITIEKSPAYITEEVEERYK 172
D 81 HPDVAANEVHFFDWEHYSQGLWYLTQMPFSSPHQLTVEKTPAYTPSPKVPERI 140
QY 173 NSSIKLLIIVREPTTRAISSDTQVLEGEKERNKTYKFEKLAIDPNTCEVNTKVKAVRTS 232
D 141 NPTIRLLIIRDFSERVLSYDQVLYNHLQKHKKYPPIEDLLMRDG--RLNLDYKALNRS 198
QY 233 IYTKHLERWLKYPFIEQFHVVDGDRLLITEPELQVKEFLNLPRISSQYNLYFNATRG 292
D 199 LYAHMLNWLRFPLGHHIVDGRLLIRDPPEIETQKVERFLKSPQINASNFYFNKTKGF 258
QY 293 YCIRFNIFNKCLAGSKGRHPEVDPSPVITKLRKFFHFNQKFIQITGRITLW 345
D 259 YCLR-DSGRDRCLHESKRAHPQVDPKLLNKLHIEFHEPNKFFELVGRITPDW 310

RESULT 6
Q9ESG5 PRELIMINARY; PRT; 311 AA.
ID Q9ESG5
AC Q9ESG5
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 3-O-sulfotransferase.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
OX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Wistar;
RA Li Z.Y., Hirayoshi K., Suzuki Y.;
RT "Expression of N-deacetylase/sulfotransferase and 3-O-sulfotransferase
RT in rat alveolar type II cells.";
RL Am. J. Physiol. 279:L292-L301(2000).
DR EMBL; AF177430; AAG09283.1; -;
DR HSSP; P52848; INST.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 311 AA; 35810 MW; ACEET78B365BECB75 CRC64;
Query Match 41.1%; Score 745.5; DB 11; Length 311;
Best Local Similarity 48.1%; Pred. No. 1.6e-51;
Matches 141; Conservative 54; Mismatches 79; Indels 19; Gaps 3;
QY 69 HEFRKGNASKEQVRLHDLV-----QOLPKAIIIGVRKGGTRALLEMLN 112
D 21 HPAAPGGLKQGLLRKVIILPDTGEGTAGNSTQOLPOTIIIGVRKGGTRALLEML 80
QY 113 HPAVVKASOEIHFFDNDENYKGGIWKYKMPSPYPOQITIEKSPAYITEEVEERYK 172
D 81 HPDVAANEVHFFDWEHYSQGLWYLTQMPFSSPHQLTVEKTPAYTPSPKVPERI 140
QY 173 NSSIKLLIIVREPTTRAISSDTQVLEGEKERNKTYKFEKLAIDPNTCEVNTKVKAVRTS 232
D 141 NPTIRLLIIRDFSERVLSYDQVLYNHLQKHKKYPPIEDLLMRDG--RLNVDYKALNRS 198


```
QY 233 IYTKHLERLWYFPIRQFHVVDGRLTEPLQELQVEKFLNLPRISSQVNLFPNATRGF 292
Db 199 LTHAEMNLWLPPLGLHIVDGRFIDRFPIQKVERFLKSLPQINASNFYFNKTGPF 258
QY 293 YCLRFNIIIFNKCLAGSKRGRHHPVDSVITKLRKFFHPFNQKPYQITGRLNW 345
Db 259 YCLR-DSGKDRCLHESKGRAPQVDPDKLLDKLHFEYFREPKNKFFKLVGRTFDW 310

RESULT 7
Q8MR67 PRELIMINARY; PRT; 605 AA.
AC Q8MR67
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE GH20068P.
GN CG15075.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burkova D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarni H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY121626; AAM51953.1; -.
DR FlyBase; FBgn0034385; CG15075.
SQ SEQUENCE 605 AA; 66437 MW; 128B17C938DA19C2 CRC64;

Query Match 40.0%; Score 726.5; DB 5; Length 605;
Best Local Similarity 41.5%; Pred. No. 1.2e-49;
Matches 152; Conservative 43; Mismatches 62; Indels 109; Gaps 4;

QY 88 QQLPRAIIIGVKGKGTALLENLHPVAVKASQETHTFDNDNKGIEWVKMPFSG 147
Db 240 RLPLQALIIIGVKGKGTALLENLHPVAVKASQETHTFDNDNKGIEWVKMPFSG 299
QY 148 PQOITIEKSPAYFITEVERIYKMSIKLLIIVREPTTIAISDYQVLE----- 198
Db 300 RGQITIEKSPFYFVSEPEVERVAMASIKLLIIVREPTTIAISDYQVLE----- 359
QY 199 ----- 198
Db 360 ARKDPSPRESSGGAGGGGAGGGGGTAAAKMPTQSLLYAKLOSAGYDNALGGSGA 419
QY 199 -KGRNKTY-----YKFEKLAIDPNTC 220
Db 420 GAKETKGTSSLSVLRQALGGGGVAGAAATTTTSPWASAAQAAKSFELAFPNHG- 478
QY 221 EVNTKYKAVRTIYTKHLERLWYFPIRQFHVVDGRLTEPLQELQVEKFLNLPRISS 280
Db 479 TVNEAYRPLISIMYHVLHRLWLEVPFPEQLLVVNGDRLTEPVSQKRIEAFLGIEHRVN 538
QY 281 QNLYFNATRGFVCLRFNIIIFNKCLAGSKRGRHHPVDSVITKLRKFFHPFNQKPYQIT 340
Db 539 SEHFVFNETKGYCYCLRYD-NGDRCLRETGRKHPHPVDPVVSRLRKKFFFAEYNQRFELVG 597
QY 341 RTLNWP 346
Db 598 EDLGNP 603

RESULT 8
Q9VMJ7 PRELIMINARY; PRT; 384 AA.
ID Q9VMJ7
QY 341 RTLNWP 346
Db 598 EDLGNP 603

Query Match 35.6%; Score 646.5; DB 5; Length 384;
Best Local Similarity 46.8%; Pred. No. 1.7e-43;
```

```
AC Q9VMJ7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE CG7890 protein (R01736P).
GN CG7890.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Adams M.D., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abil J.F., Agbayani A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
RA Balles R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burkova D., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko F., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarni H., Krommiller B., Li P., Liao G.,
RA Miranda A., Mungall C.J., Nuncio J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE003511; AAF48941.1; -.
DR EMBL; AY119100; AAM50960.1; -.
DR HSSP; P52848; 1NST.
DR FlyBase; FBgn0031005; CG7890.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
SQ SEQUENCE 384 AA; 43374 MW; D48346358EA72C59 CRC64;
```



```
Db 257 TFRNSAGLITTSWAIQIGLYAKHLEPWLHFFLQGMFLVSGERLVSDDPAGELRVQDF 316
Qy 273 LNLPRISQYNLYNATRGFYCLARNIIFNK--CLAGSKGRHPEVDPSVITKLKPFHP 330
Db 317 LGLKRIITDKHFYFNQTKGFFCLAKGSGKPHCLGKTKGRAHPTIAREVLRLQDRFVRP 376
Qy 331 FNOKFYQITGRTLNW 345
Db 377 FNRKPYQMTGRDFGW 391

RESULT 14
OQ256 PRELIMINARY; PRT; 390 AA.
AC Q9QZS6;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE D-glycosaminyl 3-O-sulfotransferase-3B.
GN H33ST3B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CS7BL/6J;
RX MEDLINE=99449300; PubMed=10520990;
RA Shukla D., Liu J., Blalock P., Shworak N.W., Bai X., Esko J.D.,
RA Cohen G.H., Eisenberg R.J., Rosenberg R.D., Spear P.G.;
RT "A novel role for 3-O-sulfated heparan sulfate in herpes simplex virus
RT 1 entry.";
RL Cell 99:13-22(1999).
DR EMBL; AF168992; AA04505.1; -.
DR HSSP; P52848; INST.
DR MGB; MG1:1333853; Hs3st3b.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 390 AA; 43326 MW; ACD0D28D66B3DDE8 CRC64;

Query Match 33.9%; Score 614.5; DB 11; Length 390;
Best Local Similarity 45.2%; Pred. No. 6.1e-41;
Matches 127; Conservative 43; Mismatches 92; Indels 19; Gaps 5;

Qy 68 LHEFRKGNASKEQVRLHDLVQQLPKAIIIGVRKGGTRALLEMLNHPAVVKASOEIHFDD 127
Db 125 ISSFFSGAGSK-----QLPQAIIGVRKGGTRALLEFLRVHPDRAVGAEPHFF- 173

Qy 128 NDENYKGIWYRKMPSPYQQITIEKSPAYFITEVPERIYKMNSSIKLLIIVREPTT 187
Db 174 -DRSYHRGLAWYRLMPRLTKGQITMEKTPSYFVTREAPARISAMSKDTKLIIVVRDPVT 232

Qy 188 RAISDYTVQLGKGRKNKYKYEKLAIDPNTCE-VATKYKAVTSIYTKHLEWLVKXFP 245
Db 233 RAISDYTVQLS-----KRPDIPSFSLFRNKSAGLIDTSWAIQIGLYAKHLEPWLHFF 288

Qy 247 IEQHVVDGRLITEPLELQVLEKFLNLPRIISQYNLYFNATRGFYCLRNIIIFNK--C 304
Db 289 LGQMLFVSGERLVSDDPAGELRVQDFGLGKRIITDKHFYFNQTKGFFCLKKAEGSGKPHC 348

Qy 305 LAGSKGRHPEVDPSVITKLKPFHPFNQKPYQITGRTLNW 345
Db 349 LGKTKGRAHPTIAREVLRLQDRFVRPNRKYFQMTGRDFGW 389

RESULT 15
Q96RX7 PRELIMINARY; PRT; 311 AA.
AC Q96RX7
AC Q96RX7;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
```

```
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Heparan sulphate D-glucosaminyl 3-O-sulfotransferase-3B like.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21096910; PubMed=11157797;
RA Daniels R.J., Peden J.F., Lloyd C., Horsley S.W., Clark K.,
RA Tufarelli C., Kearney L., Buckle V.J., Doggett N.A., Flint J.,
RA Higgs D.R.;
RT "Sequence, structure and pathology of the fully annotated terminal 2
RT Mb of the short arm of human chromosome 16.";
RL Hum. Mol. Genet. 10:339-352(2001).
DR EMBL; AE006640; AAK61299.1; -.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
KW Transferase.
SQ SEQUENCE 311 AA; 34694 MW; 6841B6151BA0DA6F CRC64;

Query Match 33.5%; Score 608; DB 4; Length 311;
Best Local Similarity 44.7%; Pred. No. 1.5e-40;
Matches 119; Conservative 46; Mismatches 81; Indels 20; Gaps 4;

Qy 88 QQLPKAIIIGVRKGGTRALLEMLNHPAVVKASOEIHFDDNENYKGIWYRKMPSPY 147
Db 57 RRFQALIVGVRKGGTRALLEFLRLHDPDVALGSEPHFF--DRCYERGLAWYRSLMPRTL 114

Qy 148 PQQITIEKSPAYFITEVPERIYKMNSSIKLLIIVREPTTRAISDYTVQLGKGRKNKY 207
Db 115 DQGITMEKTPSYFVTREAPRIHAMSPTKLIIVVRNPVTRAISDYTVQLS---KTPGL 170

Qy 208 YKFEKLAIDENTCEVNTKYKAVTSIYTKHLEWLVKXFPYEQHFVVDGRLITEPLELQ 267
Db 171 PSFRALAFRHGLGPVDFAWSAVRIGLYAQHLDHMLRYFPLSHFLFVSGERLVSDDPAGEVG 230

Qy 268 LVKFLNLPRIISQYNLYFNATRGFYCLRNIIIFNK-----CLAGSKGRHPEVDPS 319
Db 231 RVQDFGLGKRVVTDKHFYFNATKGFPCLK-----KAQGGSRPRCLGSKGRPHRPVPOA 284

Qy 320 VITKLKRFHPENOKFYQITGRTLNW 345
Db 285 VVRLLQEFYRPNRRFYQMTGRDFGW 310

Search completed: December 4, 2003, 16:44:00
Job time : 44 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 4, 2003, 16:43:06 ; Search time 29 Seconds
(without alignments)
504.812 Million cell updates/sec

Title: PCT-US03-21094-2

Perfect score: 1814

Sequence: 1 MLFQQAALRQKLLVLSLA.....FFHFNQKFVQITGRTLNWP 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5A COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/5B COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/6A COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/6B COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/PCFUS COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1646	90.7	321	4	US-09-735-935-2
2	772.5	42.6	255	4	US-09-735-935-4
3	99	5.5	803	1	US-08-062-368-4
4	95.5	5.3	1088	2	US-08-742-026-2
5	95.5	5.3	1088	2	US-08-742-026-23
6	91	5.0	1525	3	US-09-396-651B-1
7	90.5	5.0	1076	4	US-09-134-001C-4037
8	88.5	4.9	473	4	US-09-107-532A-4200
9	88.5	4.9	600	4	US-09-996-243-347
10	88	4.9	383	4	US-09-107-532A-5773
11	87.5	4.8	296	4	US-09-134-001C-2976
12	87	4.8	431	4	US-09-107-532A-7189
13	86.5	4.8	283	2	US-08-332-562A-136
14	86.5	4.8	426	4	US-09-134-001C-3346
15	86.5	4.8	467	4	US-09-134-001C-4200
16	86.5	4.8	3898	3	US-08-750-717-2
17	86	4.7	411	4	US-09-328-352-6424
18	86	4.7	757	4	US-09-107-532A-7189
19	86	4.7	1211	4	US-09-328-352-7967
20	85.5	4.7	395	4	US-09-328-352A-24415
21	85.5	4.7	426	3	US-08-676-444-40
22	85	4.7	555	4	US-09-461-325-251
23	85	4.7	1307	4	US-09-252-991A-20867
24	84	4.6	345	4	US-09-328-352-7063
25	83.5	4.6	222	4	US-09-252-991A-26656
26	83	4.6	455	4	US-09-107-532A-6755
27	83	4.6	460	4	US-09-134-001C-2979

28	83	4.6	581	4	US-09-649-747A-13	Sequence 13, Appl
29	83	4.6	581	4	US-09-649-747A-21	Sequence 21, Appl
30	83	4.6	749	1	US-08-046-508-2	Sequence 2, Appl
31	82.5	4.5	428	4	US-09-107-532A-6453	Sequence 6453, Ap
32	82.5	4.5	449	4	US-09-984-880-2	Sequence 2, Appl
33	81.5	4.5	482	4	US-09-252-991A-20129	Sequence 20129, A
34	81.5	4.5	522	4	US-09-461-325-220	Sequence 220, App
35	81.5	4.5	574	4	US-09-996-243-340	Sequence 340, App
36	81.5	4.5	584	4	US-09-198-452A-700	Sequence 700, App
37	81	4.5	302	4	US-09-107-532A-4736	Sequence 4736, Ap
38	81	4.5	631	4	US-09-134-001C-3843	Sequence 3843, Ap
39	81	4.5	666	4	US-09-134-001C-5465	Sequence 5465, Ap
40	80.5	4.4	582	2	US-08-899-244-2	Sequence 2, Appl
41	80.5	4.4	582	3	US-09-224-772-2	Sequence 2, Appl
42	80.5	4.4	582	3	US-09-227-804-2	Sequence 2, Appl
43	80.5	4.4	724	1	US-07-906-349A-5	Sequence 5, Appl
44	80.5	4.4	724	1	US-08-167-035-2	Sequence 2, Appl
45	80.5	4.4	724	1	US-08-208-887A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-735-935-2
; Sequence 2, Application US/09735935
; Patent No. 6420150

; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al

; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN

; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF

; FILE REFERENCE: CL000970
; CURRENT APPLICATION NUMBER: US/09/735,935

; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

; TYPE: PRT
; LENGTH: 321

; ORGANISM: Human
US-09-735-935-2

Query Match 90.7%; Score 1646; DB 4; Length 321;
Best Local Similarity 100.0%; Pred. No. 2.9e-164;
Matches 311; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	36	RLQPCIEGRIGGARTQAEFFLRALQKRGLLHFRKGNASKEQVRLHDLVQOLPKAII	95
Db	11	RLQPCIEGRIGGARTQAEFFLRALQKRGLLHFRKGNASKEQVRLHDLVQOLPKAII	70
QY	96	IGVRKGGTRALLEMLNLHPAVVKAQEIHFNDENYKGIWYKMKMPFSYPOQIITIEK	155
Db	71	IGVRKGGTRALLEMLNLHPAVVKAQEIHFNDENYKGIWYKMKMPFSYPOQIITIEK	130
QY	156	SPAYITIEVPRIYKMNSSIKLLIIVREPTTRAIISDYTVQLGKERKNKTYKFEKLAI	215
Db	131	SPAYITIEVPRIYKMNSSIKLLIIVREPTTRAIISDYTVQLGKERKNKTYKFEKLAI	190
QY	216	DNTCEVNTKYKAVRTSIYTKHLEWLKYPPIEQHVHVDGDLITEPLPELQVKEFLNL	275
Db	191	DNTCEVNTKYKAVRTSIYTKHLEWLKYPPIEQHVHVDGDLITEPLPELQVKEFLNL	250
QY	276	PPRISQYNLYFNATGFFYCLRFNIIFNKCLAGSKRIIHPVDPSVITLKRFFHFNQKF	335
Db	251	PPRISQYNLYFNATGFFYCLRFNIIFNKCLAGSKRIIHPVDPSVITLKRFFHFNQKF	310
QY	336	YOITGRTLNWP 346	
Db	311	YOITGRTLNWP 321	

```
RESULT 2
US-09-735-935-4
; Sequence 4, Application US/09735935
; Patent No. 6420150
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CLO00970
; CURRENT APPLICATION NUMBER: US/09/735,935
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 255
; TYPE: PRT
; ORGANISM: Human
; US-09-735-935-4

Query Match      42.6%; Score 772.5; DB 4; Length 255;
Best Local Similarity 53.1%; Pred. No. 8.2e-73;
Matches 137; Conservative 52; Mismatches 66; Indels 3; Gaps 2;

QY 88 QQLPKAIIGVRKGGTRALLEMLNLHPAVVKASQEIHFDDENYKGIWYKMKPFYS 147
Db 1 QQLPOTIIIGVRKGGTRALLEMLNLHPDVAAAEVHFDDWEHSHGLWYLSQMPFSW 60
QY 148 PQQITIEKSPAYFTEEPRIYONSSIKLLIIVREPTTRAIISDYTVQVLEKGRKNKY 207
Db 61 PHQITVEKTPAYFSPKPEPVYGMNPSIRLLILRDPSEKVLSDYTVQVFNHMQEKPY 120
QY 208 YKPEKLAIDPNTCBWNTKYKAVRTSIYTKHLERMLKYPPIQGFHVVDGDRILTEPLQ 267
Db 121 PSTIEFLVRDG--RLNDYKRAINSLTHVQNWLRFFPLHHIIVDGDRLIRDFPEIQ 178
QY 268 LVEKFLNLPRIISQNYLFNATRGFCYCLRFNIENKLAGSKGRIHPEVDPSPVITKLRF 327
Db 179 KVERFLKSLPQINASFVFNKTKGYCLR--DSGRDCLHESKGRAPQVDPKLNKLHEY 237
QY 328 FHPFNQKPYQITGRTLAW 345
Db 238 FHPNKKXFFELVGRFTDM 255

RESULT 3
US-08-062-368-4
; Sequence 4, Application US/08062368
; Patent No. 5491086
; GENERAL INFORMATION:
; APPLICANT: Gelfand, David H.
; APPLICANT: Wang, Alice M.
; TITLE OF INVENTION: Purified Thermostable Nucleic Acid
; TITLE OF INVENTION: Polymerases Enzyme From Pyrodictum Species
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/062,368
; FILING DATE: 19930514
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Sias Ph.D., Stacey R.
; REGISTRATION NUMBER: 32,630
; REFERENCE/DOCKET NUMBER: 8584
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2863
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 803 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-062-368-4

Query Match      5.5%; Score 99; DB 1; Length 803;
Best Local Similarity 18.6%; Pred. No. 0.2;
Matches 63; Conservative 44; Mismatches 106; Indels 126; Gaps 14;

QY 81 VRLHDLVQOLPK-----AIIIGVRKGGTRALLEMLNLHPAVVKASQEIHFDDNDEN 131
Db 295 VDIYDYAEEMPEIKMKTLSEVAYLGVKKSERVIVIEWRIP-----BYWD-DEK 343
QY 132 YKGIEWY-----RKMPFSYPOQIITIEKSP-----AY 159
Db 344 KRQLERYALDDVRATYGLAEKMLPEAI--QLSTVTGVPLDQVGANGVGFRLWYLMEAY 402
QY 160 FITEEPEYRIYKMSIKLLIIVREPTTRAIISDYTVQVLEKGRKNKYKF-----EKLAI 215
Db 403 DMNELVFNVRERGESYKGAWLK--PLKGVHENVVVLDFSSWYPMIKYINVGPDITVD 460
QY 216 DPNTC-----EVNTKYKAVRTSIYTKHLERMLKYPPIQGFHVVDGDRILTEPLPE 265
Db 461 DPSECKYGCYVAPGVHFRSPGCFKTVLENLLK-----R 500
QY 266 LQIVKEFLNLPRIISQNYL-----FNATRGF-----YCLRF-----N 298
Db 501 RQVKERMKPEPPDPSPEYRLYDERQKALKVLNAYSGYGMGSHARWYCKRCACAAVTAMGRN 560
QY 299 IIFN-----KCLAGSKGRIHPEVDPSPVITKLRF 327
Db 561 LILTAIEYARKLGLKVIYGTDSLFSVYDYKKEKVEKLIEF 599

RESULT 4
US-08-742-026-2
; Sequence 2, Application US/08742026
; Patent No. 585815
; GENERAL INFORMATION:
; APPLICANT: Sassanfar, Mandana
; APPLICANT: Kaufmann, Christoph
; APPLICANT: Gallant, Paul L.
; APPLICANT: Kranz, Janice E.
; APPLICANT: Houtman, Fariba
; TITLE OF INVENTION: Candida Isocetyl-trNA Synthetase
; TITLE OF INVENTION: Proteins, Nucleic Acids and Strains
; TITLE OF INVENTION: Comprising Same
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,026
; FILING DATE:
```

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI95-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1088 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-742-026-2

Query Match 5.3%; Score 95.5; DB 2; Length 1088;
Best Local Similarity 19.8%; Pred. No. 0.73;
Matches 44; Conservative 48; Mismatches 63; Indels 67; Gaps 11;

QY 123 IHFFDNDENYKGIWY-----RKMPFSYPOQITIEKSPAYFITEEVEPERIYKMNSSIKL 178
DB 731 LHFIDDLTN-----WYIRNRRRI-----KGYASDDVEDTQKGLNTLVEA 770
QY 179 LIIVREPTTTRAIISDYTQVL-EGKERKNKTYK---PEKLAIDPNTCEVNT-----KYKAVR 230
DB 771 LTLT-----SRAMAPFTPYLADGIYQRIKVFQKQEDLEKIAINPKNVDLRSVHFLSYPSVR 826
QY 231 TSIYTKHLE-----RWLYKYPFIEQHVVDGDRLLITE-----PLPELQVLVEKFLNL--- 275
DB 827 QELFDEKIEVAVARMQK-----VIDMARNIREKKMISLKTPLNELVVLVSADADLLKD 878
QY 276 -----PPRISQNYLYFNATRGFYCLRENIIFNKCLAGSK 309
DB 879 IDSLKGYISDELNRNVVITSDEAKYCVESVADWPVLGKK 920

RESULT 5
US-08-742-026-23
Sequence 23, Application US/08742026
Patent No. 5885815
GENERAL INFORMATION:
APPLICANT: Sassanfar, Mandana
APPLICANT: Kaufmann, Christoph
APPLICANT: Gallant, Paul L.
APPLICANT: Kranz, Janice E.
APPLICANT: Houtman, Fariba
TITLE OF INVENTION: Candida Isoleucyl-tRNA Synthetase
TITLE OF INVENTION: Proteins, Nucleic Acids and Strains
TITLE OF INVENTION: Comprising Same
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742.026
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Brook, David E.
REGISTRATION NUMBER: 22,592
REFERENCE/DOCKET NUMBER: CPI95-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1088 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-742-026-23

Query Match 5.3%; Score 95.5; DB 2; Length 1088;
Best Local Similarity 19.8%; Pred. No. 0.73;
Matches 44; Conservative 48; Mismatches 63; Indels 67; Gaps 11;

QY 123 IHFFDNDENYKGIWY-----RKMPFSYPOQITIEKSPAYFITEEVEPERIYKMNSSIKL 178
DB 731 LHFIDDLTN-----WYIRNRRRI-----KGYASDDVEDTQKGLNTLVEA 770
QY 179 LIIVREPTTTRAIISDYTQVL-EGKERKNKTYK---PEKLAIDPNTCEVNT-----KYKAVR 230
DB 771 LTLT-----SRAMAPFTPYLADGIYQRIKVFQKQEDLEKIAINPKNVDLRSVHFLSYPSVR 826
QY 231 TSIYTKHLE-----RWLYKYPFIEQHVVDGDRLLITE-----PLPELQVLVEKFLNL--- 275
DB 827 QELFDEKIEVAVARMQK-----VIDMARNIREKKMISLKTPLNELVVLVSADADLLKD 878
QY 276 -----PPRISQNYLYFNATRGFYCLRENIIFNKCLAGSK 309
DB 879 IDSLKGYISDELNRNVVITSDEAKYCVESVADWPVLGKK 920

RESULT 6
US-09-396-651B-1
Sequence 1, Application US/09396651B
Patent No. 6225076
GENERAL INFORMATION:
APPLICANT: Darst, Seth A
APPLICANT: Zhang, Gongyi
APPLICANT: Campbell, Elizabeth
APPLICANT: Minakin, Leonid
APPLICANT: Severinov, Konstantin
TITLE OF INVENTION: A CRYSTAL OF BACTERIAL CORE RNA POLYMERASE AND METHODS
TITLE OF INVENTION: OF USE THEREOF
FILE REFERENCE: 600-1-258
CURRENT APPLICATION NUMBER: US/09/396,651B
CURRENT FILING DATE: 1999-09-15
NUMBER OF SEQ ID NOS: 4
SOFTWARE: Patent in Ver. 2.0
SEQ ID NO 1
LENGTH: 1525
TYPE: PRT
ORGANISM: Thermus aquaticus
FEATURE:
NAME/KEY: SITE
LOCATION: (1247)
OTHER INFORMATION: Any amino acid can be at this position
US-09-396-651B-1

Query Match 5.0%; Score 91; DB 3; Length 1525;
Best Local Similarity 20.6%; Pred. No. 3.6;
Matches 41; Conservative 31; Mismatches 65; Indels 62; Gaps 8;

QY 85 DLVQQLPKAI-IIGVRKGGTRALLEMLNLHPAVVAKASQEIHFDDNDENYKGIWYKMKM 143
DB 1252 DITQGLPRVIELFEARRPKAKAVISEIDGVVRIBEGEDRLSVFVSEGFSGKE----- 1303
QY 144 PFSYPOQITIEKSPAYFITEEVEPERIYKMNSSIKL-----IIVREPTTTRAIISDYTQVL 197
DB 1304 -----YKLPDARLLVKGQDYVEAGOLTRGALDPHQLL 1337
QY 198 EGKERKNKTYKPEKLAIDPNTCEVNTKYKAVRTSYTKHLE-----RWLYKYPFIEQHV 253
DB 1338 EAKGPE-----AVERYLVD-----EIQKVVRAQGVKLDKHEIIVVRQMLAYVEVD---- 1384


```
QY 254 DGRLLTEPELPELQVKEF 272
Db 1385 PGD-----SPLEGQVLEK 1399

RESULT 7
US-09-134-001C-4037
; Sequence 4037, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4037
; LENGTH: 1076
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4037

Query Match 5.0%; Score 90.5; DB 4; Length 1076;
Best Local Similarity 21.7%; Pred. No. 2.4;
Matches 58; Conservative 40; Mismatches 86; Indels 83; Gaps 14;

QY 45 GRLGGARTQAE-FPL--RALQPK-----RGLLHEF-----RKNASKEQVRLHDLVQQ 89
Db 723 GLIGGKGRANLYGLFNKAVPENSSFGLFQFIRFIDELIDRKKDFGEENV----- 774

QY 90 LPKAIILGVRKGGTRALLEMLNH-----PAVKASQEIHFDDN-----YCK 134
Db 775 -----VGPDNDVVMWMTIHSSKGLGFPPFIYSGLSKFKNGKDLNAPVLNQOYGL 824

QY 135 GLEWVRKMPFVSPQITTEKSPAY-----FITEVPERIVKMSSIKLILIVREPTT 187
Db 825 GMDYFDVNMKMFPSLASV-----AYRAINEKELISEE-----WRLIYVA---IT 866

QY 188 RAISDYTVQLEGERKRNKTYFPEKLAIDPNTCEVNTKYKAVRT-----SIYTKHLERML 242
Db 867 RAKEQL--ILVGRVKDEKSLIKYKLAVSDTHIAVNERLTATNPVFLIVGILAKHQSPSL 924

QY 243 KYFPIEQFHVVGDRLITEPELQV 269
Db 925 ---PNDQRFERDIDQLNSEVKPRVSIV 948

RESULT 8
US-09-107-532A-4200
; Sequence 4200, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
```

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Denise
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 473 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...473
; SEQUENCE DESCRIPTION: SEQ ID NO: 4200:
US-09-107-532A-4200

Query Match 4.9%; Score 88.5; DB 4; Length 473;
Best Local Similarity 23.5%; Pred. No. 1.1;
Matches 50; Conservative 36; Mismatches 80; Indels 47; Gaps 10;

QY 108 EMLNLHPAVKASQEIHF---FDNDENYKGIWVRKMPFSPQITTEKS-----PAY 159
Db 209 EKINYLQVTKRKELFLNTFTFENLTMINPNI-----MNFEDSIEASGFRIOPEY 260

QY 160 FITEVPERIVKMSSIKLIL-----IVREPTTSAISDYTVQLEGERKRNKTYK 209
Db 261 ---ELFRLFDIIVSLILLILASFPMLITAILVKTTSFGIIVYQVTRITKQKESFYK 316

QY 210 FEKLA--IDPNTCEVNTKYKAVRTSIYTKHLERMLKYFPIEQ-FHVVDGDRLITEPEL 266
Db 317 FRMSATAEAKSGPVLAKSNDARVTPVGKFI-RAVRPELQIFNVLKGDSIVGPRP- 374

QY 267 QLVKFLNLPRIISOYNIYFNATRGFYCLRFNI 299
Db 375 -----RPFVDO-----FNEENPYIYLRHV 395
```

```
RESULT 9
US-09-996-243-347
; Sequence 347, Application US/09996243
; Patent No. 6478825
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Geritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
```

APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C13
CURRENT APPLICATION NUMBER: US/09/996,243
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090535
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090540
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090542
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24


```
; Sequence 2976, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2976
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
; US-09-134-001C-2976

Query Match      4.8%; Score 87.5; DB 4; Length 296;
Best Local Similarity 24.6%; Pred. No. 0.7;
Matches 31; Conservative 25; Mismatches 59; Indels 11; Gaps 5;

Qy 139 YRKNPFSYPOQITIE-KSPAYFITEEVEPERIYKNSSIKLLIIVREPTTTRAIISDYTVL 197
Db 54 YESKSWISAPKEITYDYKSESYKLEQNSNNFYMYDILTALYLTTPKLSHYFYQYLKL 113
Qy 198 EGKERKN-KTVYKF-EKLAIDPNTCEVNTK---YKAVRTSIYTKHLERWL---KYPI 247
Db 114 IVKHSNRYTLLKYLEFIDKNGTGSTSLVQAMNENQYLOYNQNLPLISMYQM 173
Qy 248 BQHVY 253
Db 174 HAFHLV 179

RESULT 12
US-09-107-532A-5796
; Sequence 5796, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
```

```
; INFORMATION FOR SEQ ID NO: 5796:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...431
; SEQUENCE DESCRIPTION: SEQ ID NO: 5796:
US-09-107-532A-5796

Query Match      4.8%; Score 87; DB 4; Length 431;
Best Local Similarity 22.2%; Pred. No. 1.4;
Matches 51; Conservative 49; Mismatches 98; Indels 32; Gaps 14;

Qy 124 HFF--DNDENVGKG-IWYRKMPFSYPOQITIEKSPAYFITEEVEPERIYKNSSIKLLI 180
Db 152 HFFADDDGEQSFSETFIQLKQKRD---PASTIEILGDSRTLFFELPELLFREBQ---IL 204
Qy 181 IVREPTTTRAI--SDYTOVLEKGRKNKYKFEKLAIDPNTCE---VNTKYKAVRTSIY 234
Db 205 PIKKNVLAIRLSCNQITKRWRIEETQDYRKLTIUTEKLDLSLFWLNSEFSGNKS 264
Qy 235 TKHL--ERWLKYPFLEQFHVVDGRLITEPLPELQ-LVEKFLNLPFRISQYNLYFNATRGF 292
Db 265 TKVLYQEWLRY--VIGIH---DLVEEVRYYYQETLEKFPNSKDLVQPLAIQALAH 318
Qy 293 YCLRFNIIFNKLAGSKG--RIHPEVDPSPVITKURKFFHPFNQKPYQITG 340
Db 319 PAVWGMGLLTKLQ-ERGYLDVHD---PEV--KIVFFFRYEHQBARLVG 362

RESULT 13
US-08-332-562A-136
; Sequence 136, Application US/08332562A
; Patent No. 5985599
; GENERAL INFORMATION:
; APPLICANT: McKENZIE, Ian F.C.
; APPLICANT: HOGARTH, Mark P.
; APPLICANT: HIBBS, Margaret L.
; APPLICANT: SCOTT, Bernadette M.
; APPLICANT: BONADONNA, Lisa
; APPLICANT: HULETT, Mark D.
; TITLE OF INVENTION: FC RECEPTOR FOR IMMUNOGLOBULIN
; NUMBER OF SEQUENCES: 136
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,562A
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/896,457
; FILING DATE: 27-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 54270/119/GRHA
; TELECOMMUNICATION INFORMATION:
```

TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 1
SEQUENCE CHARACTERISTICS:
LENGTH: 283 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-332-562A-136

Query Match	4.8%;	Score	86.5;	DB	2;	Length	283;		
Best Local Similarity	23.9%;	Pred. No.	0.84;						
Matches	53;	Conservative	31;	Mismatches	85;	Indels	53;	Gaps	10;
Qy	45	GRLGGAQTQAEFFLRLAQFKRGLLHHPKGNASKEQVRLHD-----LVQOLPKAI	94						
Db	74	GRSIRSQVQASYFKATVNDSG---EYR---CQEQTRLSDPVDLGVISDWLLLTQFLV	127						
Qy	95	IIGVRKGGTRALLEMLHPAVKVASQEIHFDDNDENYGKIEWYRKMPFSYQQ----	150						
Db	128	FL---EGETIT---LRCHSRWNKLLNRISPFHNE---KSVRYHYSSNFSFIPKANHSH	176						
Qy	151	-----ITIEKSPAFITEEVEPERIYKMNSSIKLILIVEPTTRALSDYTVQLEG	199						
Db	177	SGDYICKSGSLGRUHQKSPVITVQGP-----KSRSLPVLITVAAVTGIAVAIVILL--	230						
Qy	200	KERKNKTYIKPEKLAIDPNTCEVNTKYKAVRTSYT--KHLE	239						
Db	231	---VSLVYLKKQVDPNDPPLQEEAAKTAENTIYISLLKHEP	269						

RESULT 14
US-09-134-001C-3346
; Sequence 3346, Application US/09134001C
; Patent NO. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3346
; LENGTH: 426
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3346

```
Query Match          4.8%; Score 86.5; DB 4; Length 426;  
Best Local Similarity 21.9%; Pred. No. 1.6;  
Matches      64; Conservative    38; Mismatches   89; Indels 101; Gaps   14;
```

Qy	59	RALQFKGLLHFRKGNASKEQVRLHDL-----VQOLPKAIIIVRGKGTRALLEM	109
Dd	8	KALLFWQLI-----SNEERTVMTHITIVGLNGYIDELPLGIYFLEX-ESKVART	61
Qy	110	LNLHPAVVKASQEIHFFDNDENYGGKIETWYKKMPFYQPQQITTEKSPAFVITEVEPERI	169
Dd	62	LN-HPVINTLKKEIEFSFSIY-----EAHDRFDVDYEAI	96
Qy	170	YKMSSIKLL-----II-----VRBPTTAISDYT-----OVLEGKERKNKYK	209
Dd	97	--VTSLLELAQSIEDIVAVPGCHPRVATTIVTKLEYSHFNKDLSVKVLGGSFIDDI---	151
Qy	210	PEKLADPN-----TCBVNTKYKAVRTSIYTKHLERWLKYPPIEQ----	252
Dd	152	FEADVDPNPDGTILLDTSLKESALNVRTNFVIQTVSVMIAADKLTLMEVPDDEPNVK	211

Qy 253 -----VDGRLLTEPLPELQLVEKFLNLPRIQQYNLYFNATRGFYCLRFN 298
 || | | | | | | | | | |
Dd 212 IITGSHSDGAHVIECPLEY-----IDRYDDYFNNLTSLFIPIKIN 250

```

RESULT 15
US-09-134-001C-4200
; Sequence 4200, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4200
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4200

```

Query Match	4.8%;	Score	86.5;	DB	4;	Length	467;
Best Local Similarity	22.9%;	Pred. No.	1.8;				
Matches	52;	Conservative	31;	Mismatches	73;	Indels	71;
						Gaps	13;
QY	52	TOAEF-PLRALQPKRGLLHEFRKGNASKQVRUHLDLVQQ-----LPKAILIIGVRKG-G	102				
Db	269	SQAQYDAILDMNNGIKHPF-----ETVKAELSTQNAEDDSLFPVLLVVEQKGG	321				
QY	103	TRA-----LLEMLNLHPAVKVASQBIHFDFDNDENYK-----GIEWRKKMKPFSPQOITTI	153				
Db	322	LKAAEELLOKTNANPVITSGHPD---FTEDMLYKVLPNESGALYREG-----	367				
QY	154	EKSPAFIIEEVEPERIYKNNSSIKLLIIVREPITRAISDYTVQLEKGE-----	201				
Db	369	-----FIHLMP-----KKDKSVLIDNFNRINDVIDPQTVINLVEGYEVTLPYRNKDGSM	417				
QY	202	---RKNK-TYYPEKLAIDPNTCEVNTKYKAVRTSIYTKHLERWKY	244				
Db	418	IKWSKNKDSFYHP-----NPNWHIVGTTVDSIB-KIKOKYSSOFLKY	458				

Search completed: December 4, 2003, 16:46:41
Job time : 31 secs